Mock Galaxy Catalogues And Their Application To Future Galaxy Surveys

MERSON, ALEXANDER,IAN

Mock Galaxy Catalogues And Their Application To Future Galaxy Surveys

Alexander Ian Merson

Abstract

We present a method for constructing end-to-end mock galaxy catalogues using a semi-analytical model of galaxy formation, applied to the halo merger trees extracted from a cosmological N-body simulation. These mocks are *lightcone* catalogues, which incorporate the evolution of galaxy properties with cosmic time. Interpolation is used to determine the epoch at which a galaxy will appear in the past lightcone of the observer. We discuss several applications of mock catalogues. Firstly, we consider the effectiveness of the BzK colour selection technique. The mock catalogue predictions are in reasonable agreement with the observed number counts of BzK galaxies. We predict that over 75 per cent of the model galaxies with $K_{AB} \leqslant 23$, and redshift $1.4 < z < 2.5$, are selected by the BzK technique. Interloper galaxies, outside the target redshift range, are predicted to dominate bright samples of BzK galaxies (i.e. with $K_{AB} \leqslant 21$). Fainter K-band cuts are necessary to reduce the predicted interloper fraction. Secondly, we use a mock catalogue to calibrate a galaxy group-finding algorithm, via an objective method based upon the recovery of the distributions of several, easily measurable group properties. We find that it is extremely difficult to determine unique values for the linking lengths by minimising the χ^2 statistic for individual properties, and that it is necessary to combine χ^2 for more than one group property to reduce the parameter space. However, based upon our calibration, we conclude that the optimal linking lengths depend upon the multiplicity of the groups and the group property that one wishes to recover. For our final application, we use a lightcone catalogue to estimate the cosmology-independent angular correlation function, $\omega(\theta)$, for samples of galaxies, selected in bins of apparent magnitude, in a thin redshift slice comparable to the size of photometric redshift errors. We compare our estimates of $\omega(\theta)$ with the `GALFORM` predictions of the 3-dimensional real-space and redshift-space correlation functions. The amplitude of the real-space and redshift-space correlation functions display a trend with increasing luminosity. However, this trend is less clear in $\omega(\theta)$ due to noisy estimates for the brightest two apparent magnitudes bins.

Mock Galaxy Catalogues And Their Application To Future Galaxy Surveys

by

Alexander Ian Merson

A thesis submitted to Durham University
in accordance with the regulations for
admittance to the Degree of Doctor of Philosophy.

Department of Physics
Durham University
September 2012

Contents

List of Figures

List of Tables

Declaration

The work described in this thesis was undertaken between 2008 and 2012 while the author was a research student under the supervision of Prof. Carlton Baugh in the Department of Physics at Durham University. This work has not been submitted for any other degree at Durham University or any other University.

Chapter 3 and 4 of this thesis have been submitted in the form of a paper

- Merson, Alexander I.; Baugh, Carlton M.; Helly, John C.; Gonzalez-Perez, Violeta; Cole, Shaun; Bielby, Richard; Norberg, Peder; Frenk, Carlos S.; Benson, Andrew J.; Bower, Richard G.; Lacey, Cedric G.; Lagos, Claudia del P., "Lightcone mock catalogues from semi-analytic models of galaxy formation - I. Construction and application to the BzK colour selection", MNRAS in press, astro-ph/1206.4049

Lightcone mock catalogues constructed using the method discussed in Chapter 3 have been used in the following papers:

- Robotham, A. S. G.; Norberg, P.; Driver, S. P.; Baldry, I. K.; Bamford, S. P.; Hopkins, A. M.; Liske, J.; Loveday, J.; Merson, A.; Peacock, J. A.; Brough, S.; Cameron, E.; Conselice, C. J.; Croom, S. M.; Frenk, C. S.; Gunawardhana, M.; Hill, D. T.; Jones, D. H.; Kelvin, L. S.; Kuijken, K.; Nichol, R. C.; Parkinson, H. R.; Pimbblet, K. A.; Phillipps, S.; Popescu, C. C.; Prescott, M.; Sharp, R. G.; Sutherland, W. J.; Taylor, E. N.; Thomas, D.; Tuffs, R. J.; van Kampen, E.; Wijesinghe, D., "Galaxy and Mass Assembly (GAMA): the GAMA galaxy group catalogue (G^3Cv1)", MNRAS, **416**, pp. 2640-2668

- Laureijs, R.; Amiaux, J.; Arduini, S., *et al.*, "Euclid Definition Study Report", astro-ph/1110.3193

- Alpaslan, Mehmet; Robotham, Aaron S. G.; Driver, Simon; Norberg, Peder; Peacock, John A.; Baldry, Ivan; Bland-Hawthorn, Joss; Brough, Sarah; Hopkins, Andrew M.; Kelvin, Lee S.; Liske, Jochen; Loveday, Jon; Merson, Alexander; Nichol, Robert C.; Pimbblet, Kevin, "Galaxy And Mass Assembly (GAMA): estimating galaxy group masses via caustic analysis", MNRAS in press, astro-ph/1204.0510

Early versions of the groups catalogue discussed in Chapter 5 have been used in the following paper:

- Magoulas, Christina; Springob, Christopher M.; Colless, Matthew; Jones, D. Heath; Campbell, Lachlan A.; Lucey, John R.; Mould, Jeremy; Jarrett, Tom; Merson, Alex; Brough, Sarah, "The 6dF Galaxy Survey: The Near-Infrared Fundamental Plane of Early-Type Galaxies", MNRAS in press, astro-ph/1206.0385

Acknowledgments

First of all I would like to thank Carlton Baugh for being the best supervisor that anybody could ask for. I am indebted to you, Carlton, for your guidance, your advice, your support, your knowledge and your patience that have made the last four years working with you such a privilege.

Special thanks must also go to Peder Norberg for his role as an additional, unofficial supervisor and mentor and for the many invaluable and insightful discussions that, over the past couple of years, have greatly helped to develop my knowledge and my understanding.

Much of this work has been carried out on The Cosmology Machine "COSMA" in Durham. I am eternally grateful to Lydia Heck, John Helly, Peter Draper and Alan Lotts for their persistent help whenever I encountered computing problems (especially for the computing problems that I caused myself)!

I have been lucky enough to have made so many good friends during my time in Durham and to give every one of them the thanks they deserve would need a thesis in itself! I feel, however, that I must draw attention to a few names.

A huge thank you must go to Claudia Lagos, Nikos Nikoloudakis, Violeta Gonzalez-Perez, Michelle Furlong, Gabriel Altay, Carla Handley, Alvaro Orsi, Nikos Fanidakis, David Murphy, Julie Wardlow, Floyd Jackson, Alice Mills and Nicolas Tejos. Thank you for all of the good times shared, all of your support in the bad times and for all of the amazing experiences that we've had in these past few years! You are all true friends who I will never forget. To add to this list, I would also like to thank Alex Machin, Peter Holland, James Brzezicki, Christopher Robertson, Christopher Phang, Tomasz Dunn, Richard Smith, David Dodd, Thomas Tomlinson, Katie Wrathmell and Lynnette Haddock. Thank you also to all of the wonderful friends that I have made at Christchurch Durham over the past year and whilst playing Ultimate with Ashes, DUF and The Brown. You are all awesome.

Finally, I would like to thank my family. Firstly, thank you Pat, Anne and Harold for your continued support and encouragement that you have shown me throughout my education. And last, but by no means least, I owe a great deal to my parents, Anne and Iain. Mum and Dad, thank you for absolutely everything that you have done for me, for all of the sacrifices that you have made and for your never-ending love and encouragement.

This work was funded by the Sciences and Technology Facilities Council.

Chapter 1 | *Introduction*

1.1 The dark Universe

In recent decades, a vast number of observational and simulation efforts have greatly enhanced our insight into the cosmic history of the Universe and allowed us to make fundamental predictions regarding its nature. For example, precision measurements of temperature fluctuations in the *cosmic microwave background* (CMB) have been made by the COsmic Background Explorer (COBE) satellite, followed by measurements of the acoustic peaks in the power spectrum of these fluctuations by experiments such as the Balloon Observations Of Millimeter Extragalactic Radiation and Geomagnetics (BOOMERanG) and Degree Angular Scale Interferometer (DASI) experiments and, notably, the Wilkinson Microwave Anisotropy Probe (WMAP) satellite. These temperature fluctuations, believed to correspond to primordial density fluctuations in the early Universe, and the location of their acoustic peaks have enabled us to place incredibly precise constraints regarding our cosmological model for the energy constituents of the Universe as well as the growth of these fluctuations into the large-scale structure that we observe today (Smoot et al., 1992; Gorski et al., 1994; Bennett et al., 1996; de Bernardis et al., 2000; Halverson et al., 2002; Spergel et al., 2003; Dunkley et al., 2009; Komatsu et al., 2011).

The wealth of observational evidence attained over the past few decades strongly favours a Universe in which in excess of 80 per cent of matter is made up of some exotic form of matter known as *dark matter*. The concept of dark matter stems from work by Oort (1932), who measured the motions of stars in the Milky Way, and Zwicky (1933), who measured the orbits of galaxies in large clusters such as the Coma cluster. Due to the observed motion of these systems both Oort and Zwicky postulated the existence of some extra hidden form of matter. These dynamical arguments centre around the question of whether the visible matter in these systems is sufficient to provide the necessary gravitational force to keep the stars, or galaxies, in orbit. A well known example is that of galaxy rotation curves. Due to the gravitational influence of visible matter alone, which is centrally concentrated, the rotational velocity of the stars in the outskirts of a galaxy

should decrease in proportion to the square root of the radial distance from the centre of the galaxy. However, galaxy rotation curves have been observed to remain flat out to large radial distances, suggesting the presence of extra matter that is not visible but contributes to the gravitational potential (e.g. Rubin et al., 1978, 1980). Initially, this hidden matter was attributed to extremely faint compact objects, such as black holes or brown dwarfs. However, there has so far been insufficient micro-lensing signatures of massive astrophysical compact halo objects (MACHOs) to support the idea that they are responsible for the missing mass (e.g. Alcock et al., 2000). In contrast, a combination of measurements from various cosmological probes, including supernovae distances, the CMB and the observed clustering of galaxies, suggests that all matter in the Universe contributes approximately 25-30 per cent of the total energy budget, whilst additional evidence from primordial gas abundances and Big Bang nucleosynthesis suggests that the contribution from baryons alone is as little as 4 per cent. Analysis of the 7-year WMAP data by Komatsu et al. (2011), in conjunction with other data such as galaxy clustering measurements, places the total matter density of the Universe at $\Omega_{\rm M} \sim 0.27$ and the baryon matter density at $\Omega_{\rm b} \sim 0.045$, where $\Omega_{\rm M}$ and $\Omega_{\rm b}$ express the density contributions of the two components relative to the critical density of the Universe, given by

$$\rho_{\rm crit} = \frac{3H_0^2}{8\pi G}, \tag{1.1}$$

where the Hubble parameter is given by $H_0 = 100h\,{\rm km\,s^{-1}\,Mpc^{-1}}$ and $h \sim 0.70$. Since $\Omega_{\rm b} \ll \Omega_{\rm M}$, the general consensus, therefore, is that dark matter is non-baryonic.

Various particles, with vastly different masses, have been suggested as candidates for dark matter. However, evidence from the cosmological probes mentioned above currently favours a particle that is 'cold', that is a particle that becomes non-relativistic and decouples at early times[1]. As a result, possible candidates for *cold dark matter* (CDM) include exotic elementary particles such as axions or *weakly-interacting massive particles*[2]. Overall, very little is currently known of the nature of dark matter beyond that it is thought to interact with normal, baryonic matter solely through its gravitational influence.

[1]The decoupling of a particle is the removal of a particle from thermal equilibrium due to cosmological expansion. A particle is said to have decoupled once its interaction timescale exceeds the expansion timescale, set by the local Hubble time. After this point the abundance of that particle remains 'frozen'.

[2]Some observational evidence, such as the observed number of Milky Way satellites, as well as evidence for cored-density profiles of dark matter halos, is in favour of dark matter being made up of a less massive particle, such as a sterile neutrino, although various astrophysical solutions have also been proposed for these problems. Since these particles remain relativistic until later times, this form of dark matter is commonly referred to as *warm dark matter*.

Cosmological inflation, proposed by Guth (1981), is able to explain the observation that the Universe is flat with zero curvature, i.e. that the Universe has a mean density equal to the critical density, such that $\Omega_{\rm tot} = 1$, and a curvature $\Omega_k = 0$. However, as we have seen, $\Omega_{\rm M} \ll 1$, implying the existence of at least one other component to the energy density. The remaining three quarters of the energy budget of the Universe is believed to be made up of an even stranger component known as *dark energy*. This phenomenon is considered to be responsible for driving the late time accelerated expansion of the Universe that has been observed through measurements of the brightness of several hundred Type Ia supernovae (Riess et al., 1998; Perlmutter et al., 1999). These explosions of white dwarf stars (that have accreted sufficient mass such that their collapse can no longer be withheld by electron degeneracy pressure) have a known intrinsic luminosity and so can be used as 'standard candles'. Therefore, observations indicating that these supernovae appear to be fainter than expected suggested that the supernovae were further away than they were thought to be and, consequently, that the expansion of the Universe is accelerating.

The equation of state of the Universe describing the pressure, P, as a function of density, ρ, can be parameterised in terms of w,

$$P = w\rho c^2 \propto w a^{-3(1+w)} c^2 \tag{1.2}$$

where a is the expansion factor. Typically, the influence of dark energy is described using a *cosmological constant* Λ, which has a constant energy density of $\rho_\Lambda \equiv \Lambda/(8\pi G)$ that is independent of redshift. For a universe dominated by a cosmological constant, the equation of state parameter must be $w_{\rm DE} = -1$. Thus, the currently adopted standard cosmological model consisting of a flat universe dominated by a cosmological constant and with the majority of matter in the form of cold dark matter is known as the ΛCDM model. For a flat universe, we therefore have that the energy density due to the cosmological constant $\Omega_\Lambda = \Omega - \Omega_{\rm M} \simeq 0.73$. One way in which the ΛCDM cosmology can be ruled out is if $w_{\rm DE}$ is found to deviate from minus unity at any redshift[3]. Testing if $w_{\rm DE} = -1$ therefore forms one of the big questions of modern cosmology and a key goal for the next generation of galaxy surveys. So far, however, all observational measurements of $w_{\rm DE}$ are consistent with the ΛCDM model.

[3]Many other cosmological models with a late time accelerated expansion, caused by some phenomenon other than a cosmological constant, are possible for $-1 < w < -1/3$.

1.2 Hierarchical structure formation

Within the ΛCDM cosmology, the large-scale structure that we see in the present day Universe is believed to have formed from the growth of quantum fluctuations in the matter density field of the early Universe. The initial step is thought to have occurred as a result of cosmological inflation, during which the Universe underwent exponential expansion and these quantum fluctuations were transformed into macroscopic adiabatic perturbations. While the perturbations are small in amplitude, with density contrast $\delta = (\rho - \bar{\rho})/\bar{\rho} \ll 1$, where $\bar{\rho}$ is the mean density of the Universe, they grow through gravitational collapse according to linear theory, i.e.

$$\delta(\mathbf{x}, t) = \delta(\mathbf{x}, t_0)\frac{D(t)}{D(t_0)} \tag{1.3}$$

where $D(t)$ is the linear growth factor and $\mathbf{x}$ is the co-moving location of the perturbation (Peacock, 1999). Once the perturbation reaches a critical density, $\delta(\mathbf{x}, t) = \delta_{\rm c}(\mathbf{x}, t)$ (where, using the spherical collapse model, $\delta_{\rm c}(\mathbf{x}, t) \simeq 1.68$ for ΛCDM), the growth of the perturbation is halted and the perturbation begins to undergo non-linear gravitational collapse to form a virialised dark matter structure, known as a dark matter halo. In the theory of *hierarchical structure formation*, it is the smallest (lowest mass) halos that are the first to form. More massive halos form by continual accretion of material or by subsequent mergers of pairs of less massive halos (White & Rees, 1978; White & Frenk, 1991; Cole, 1991). The infall of baryons into the gravitational potential wells of dark matter halos is thought to be the first stage of galaxy formation. Upon infall, the baryons are generally assumed to be shock heated to the virial temperature of the halo. As the baryons cool, they collapse towards the centre of the halo where they form a rotationally supported disk of cold gas as a result of angular momentum conservation. We discuss the formation and evolution dark matter halos and the galaxies that they host in more detail in Chapter 2, in which we introduce the galaxy formation model that we employ throughout this work.

1.3 Galaxy surveys

Overwhelming support for the ΛCDM model has come additionally from *galaxy surveys*; extensive observational campaigns with the broad aim of measuring the positions and properties of many hundreds of thousands of galaxies and with the goal of understanding how galaxies trace out the large-scale structure of the Universe.

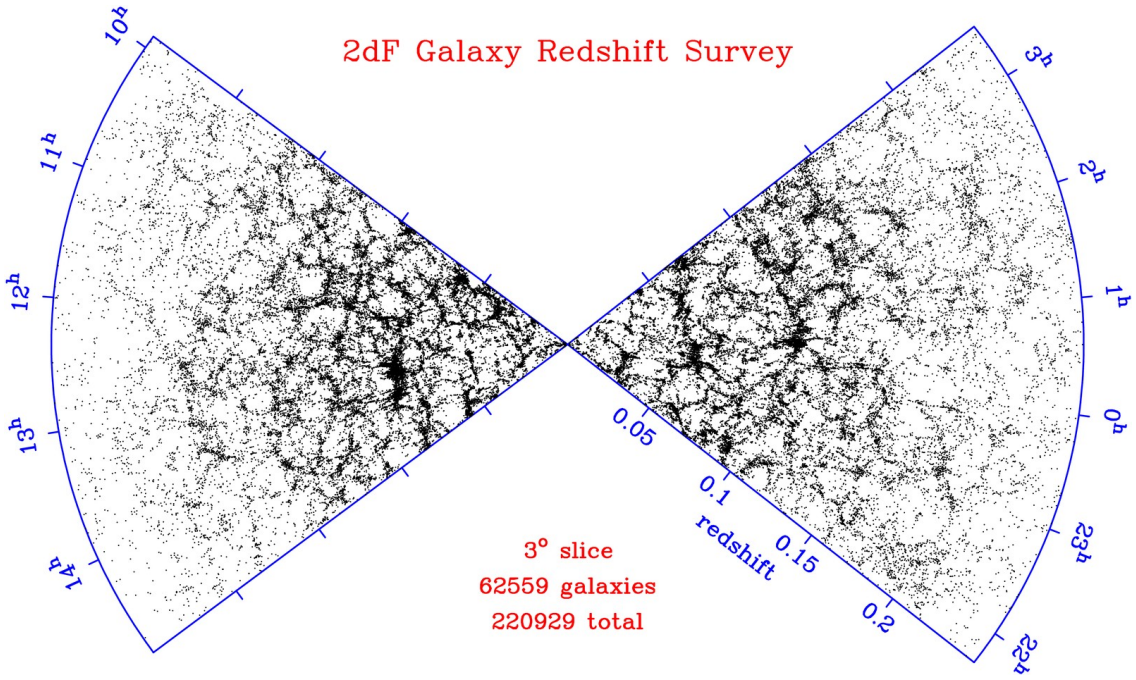

Figure 1.1: A 3° slice in declination of the Two-degree Field Galaxy Redshift Survey (2dFGRS, Colless et al., 2001), highlighting the cosmic web and large-scale structure of the Universe that can be probed with galaxy redshift surveys. Image credit: 2dFGRS Team.

1.3.1 Statistics of the galaxy population

Galaxy surveys have been invaluable for developing our understanding of the processes inherent in galaxy and structure formation, mainly due to the huge number of galaxies that they can contain and the large volumes covered. For instance, in recent years the Two-degree Field Galaxy Redshift Survey (2dFGRS Colless et al., 2001, 2003), as shown in Figure 1.1, has measured reliable positions for over 250,000 galaxies and the Sloan Digital Sky Survey[4] (SDSS, York et al., 2000; Abazajian et al., 2009) has measured the positions of approximately 930,000 galaxies, with photometric information available for many hundreds of thousands more objects. The huge size of datasets such as these, allows one to begin to explore the statistics of the population of galaxies that describe how galaxies are distributed with respect to their properties.

Amongst the simplest statistics to measure is the number counts of galaxies per unit apparent magnitude (in a particular waveband) per unit area on the sky. Galaxy number counts are able to provide some insight into the evolution of the galaxy population with redshift (e.g. Ferguson et al., 2000). However, care must be taken since the apparent

[4]We refer here to SDSS Data Release 7, which is the final data release of the SDSS-II survey.

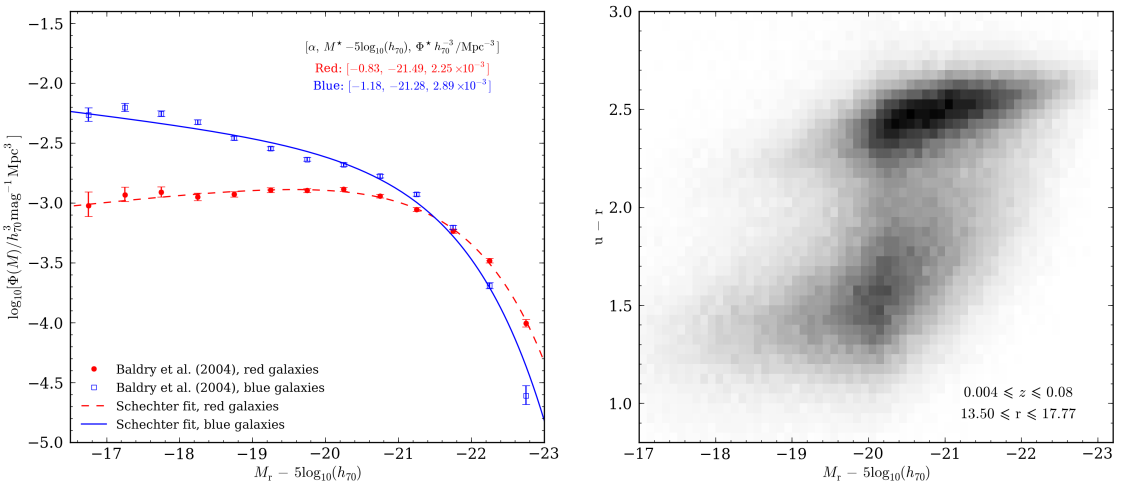

Figure 1.2: Left: Galaxy luminosity functions for red and blue selected galaxies in the SDSS. Data and Schechter fits taken from Baldry et al. (2004). Right: The local u − r colour-magnitude relation for a sample of approximately 260,000 galaxies in the SDSS. Here, the shading correlates with galaxy number density.

magnitude of a galaxy depends on both the intrinsic luminosity and the redshift of the galaxy, and so a change in the number counts could equally be caused by a change in the number of intrinsically faint objects at low redshift or in the number of intrinsically bright objects at high redshift. Additionally, it is not clear from number counts alone whether a change in the counts is due to a change in the number density of the population or a change in the luminosities of the galaxies.

Alternatively, one of the most fundamental and robust statistics of the galaxy population is the galaxy luminosity function, $\Phi(L)$, which defines the number density of galaxies with luminosities in the range L to $L+\mathrm{d}L$. As shown in Figure 1.2, the galaxy luminosity function has a characteristic shape that is generally described by a power law at faint luminosities and an exponential fall-off at bright luminosities. Recent galaxy formation models (see Section 1.4.2) have demonstrated that the faint-end slope is sensitive to the strength of feedback from supernovae in low mass halos, whilst the bright-end fall-off can be successfully reproduced by invoking feedback from *active galactic nuclei* (AGN) to quench star formation in massive halos (Benson et al., 2003; Croton et al., 2006; Bower et al., 2006). Analytically, the shape of the luminosity function can often be adequately fit by a Schechter (1976) function,

$$\Phi(M)\mathrm{d}M = 0.4\ln(10)\Phi_\star \frac{\left(10^{0.4(M_\star - M)}\right)^{\alpha+1}}{\exp\left(10^{0.4(M_\star - M)}\right)}\mathrm{d}M, \tag{1.4}$$

where α, $M^\star$ and $\Phi_\star$ are known as the Schechter parameters. The dimensionless parame-

ter α governs the faint-end slope, $M^\star$ is a characteristic magnitude and $\Phi_\star$ sets the overall normalisation. The Schechter function is often useful for quantifying evolution and variation of the luminosity function between different galaxy types. For example, we can see from Figure 1.2 that red galaxies, typically have brighter characteristic magnitudes and shallower faint end slopes than blue galaxies, suggesting that the fraction of red galaxies decreases with decreasing luminosity (e.g. Blanton et al., 2001; Madgwick et al., 2002; Baldry et al., 2004). In the 2dFGRS, Croton et al. (2005) see a similar difference between the luminosity functions of early-type galaxies and late-type galaxies, with the luminosity function of late-type galaxies resembling that of blue galaxies and the luminosity function of early-type galaxies resembling that of red galaxies. Additionally, there is evidence that the luminosity function of galaxies varies with the mass of the dark matter halo hosting the galaxies, with the galaxies in more massive halos (corresponding to galaxy groups and clusters) displaying luminosity functions with faint-end slopes that are steeper than that of field galaxies (e.g. De Propris et al., 2003; Blanton et al., 2005a; Baldry et al., 2008; Robotham et al., 2010).

The galaxy stellar mass function, which describes the number density of galaxies with stellar masses in the range M to $M + \mathrm{d}M$, can also be described by a Schechter function that qualitatively displays similar variation with galaxy type and environment to that seen in the galaxy luminosity function. Compared to the galaxy luminosity function, the galaxy stellar mass function should be a more fundamental statistic of the galaxy population, however there is greater uncertainty due to the transformation of galaxy luminosities into galaxy stellar masses.

An important result from the SDSS was the observation that the galaxy colour distribution is bimodal (e.g. Kauffmann et al., 2003b; Baldry et al., 2004; Bell et al., 2006; Brammer et al., 2009, see also González et al., 2009). The SDSS colour-magnitude relation, shown in Figure 1.2, displays two populations of galaxies: a 'red sequence' of red galaxies extending to very bright magnitudes, and a 'blue cloud' of blue galaxies, which dominates the galaxy population at fainter magnitudes. Measurements of galaxy colours in other galaxy surveys suggest that this bimodality was in place as early as $z \simeq 1$ (Bell et al., 2004; Willmer et al., 2006; Cooper et al., 2007). Observational evidence suggests that the luminosity function has undergone significant evolution over the past $\sim 8\,\mathrm{Gyr}$. However, whilst the total stellar mass density in the red population has approximately doubled since $z \sim 1$, that of the blue population has changed very little (e.g. Wolf et al., 2003; Borch et al., 2006; Faber et al., 2007). The interpretation of this result is that,

despite continued star formation occurring in the blue population, some physical process has quenched the star formation activity of a significant fraction of blue galaxies, thus moving them out of the blue cloud and onto the red sequence. With the aid of galaxy formation models (see Section 1.4.2), this effect has been attributed to the influence of AGN feedback (e.g. Bower et al., 2006). There is still, however, much debate as to the evolutionary path taken by the majority of the most massive red galaxies (e.g. De Propris et al., 2007; Faber et al., 2007; Lin et al., 2008; Tran et al., 2008; Bundy et al., 2009; de Ravel et al., 2009; Hopkins et al., 2010; Lin et al., 2010; de Ravel et al., 2011; Lambas et al., 2012).

Although we shall not discuss them here, the datasets available from galaxy surveys have helped constrain many other statistics describing the properties of the galaxy population, including the sizes of galaxies (e.g. Shen et al., 2003), the mass-metallicity relation (e.g. Gallazzi et al., 2005; Erb et al., 2006; Ellison et al., 2008; Foster et al., 2012), the fundamental plane of early-type galaxies (e.g. Bernardi et al., 2003; Magoulas et al., 2012) and the HI mass functions (e.g. Zwaan et al., 2005).

1.3.2 The spatial distribution of galaxies

Quantifying the spatial distribution of galaxies and relating this to the distribution of the underlying mass is essential for helping measure the total matter distribution, which is highly cosmology-dependent, as well as furthering our understanding of how the physical processes that play a role in galaxy formation lead to the properties of galaxies being correlated with the mass of the dark matter halos hosting them.

As revealed by galaxy surveys such as the 2dFGRS and the SDSS, the spatial distribution of galaxies, as shown in Figure 1.1, is not uniform, but instead displays a highly filamentary structure. If we assume that all galaxies reside inside virialised dark matter structures known as halos, then the large-scale distribution of galaxies can be thought of as tracing the large-scale distribution of dark matter halos (which in turn trace the matter density field). This large-scale distribution of matter is known colloquially as the *cosmic web*. In this picture, the nodes connecting the filaments and sheets are believed to correspond to the most massive dark matter structures, hosting large clusters of galaxies, whilst in between can be found large voids; regions of up to $\sim 100\,\mathrm{Mpc}$ across that contain very few, if any, galaxies.

One of the most important statistics used to quantify the galaxy and matter distribution is the *two-point correlation function* (also known as the auto-correlation function, or

simply the correlation function). This statistic describes the clustering of matter by considering the moments of the probability distribution function of the cosmic density field, $\delta(\mathbf{x})$. The two-point correlation function, $\xi(\mathbf{r})$, corresponds to the covariance between two points in the matter density field, $\mathbf{x}$ and $\mathbf{x}+\mathbf{r}$, separated by a vector distance $\mathbf{r}$, i.e.

$$\xi(\mathbf{r}) \equiv \langle \delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r}) \rangle. \tag{1.5}$$

It is often instructive to think of the two-point correlation function as the excess probability, $\mathrm{d}P$, of finding a pair of objects at a given separation, r, relative to that expected for a random distribution,

$$\mathrm{d}P = n_0^2 \left[1 + \xi(r)\right] \mathrm{d}V_1 \mathrm{d}V_2, \tag{1.6}$$

where n_0 is the number density of the population of objects and $\mathrm{d}V_1$ and $\mathrm{d}V_2$ are the volume elements inside which each of the pair of objects can be found. Although $\xi(r)$ only considers the separations between pairs of objects, the above definition can be easily expanded to consider separations between n points (also known as the n-point correlation function, $\xi^{(\mathrm{n})}(r)$).

The ease with which the two-point correlation function can be measured means that it has been estimated using specifically selected galaxy samples from various galaxy surveys (e.g. Hawkins et al., 2003; Coil et al., 2004a; Ross et al., 2007; Cabré & Gaztañaga, 2009a,b; Sánchez et al., 2009; Ross et al., 2010). However, galaxy surveys are based in redshift-space, where peculiar velocities and bulk flows distort the measured positions of galaxies (Kaiser, 1987). These *redshift-space distortions* mean that the correlation function measured in redshift space, differs from that measured in real-space. However, the projected two-point correlation function, where the galaxy separations assumed correspond to projected distances on a plane perpendicular to the line-of-sight (and therefore not affected by redshift-space distortions), can be used to translate between the real-space and redshift-space correlation functions (e.g. see Mo et al., 2010).

Clustering measurements in galaxy surveys are often made using specially selected samples of galaxies, such as Luminous Red Galaxies (LRGs), which are amongst the most luminous galaxies that are clearly visible over the volumes mapped by current galaxy surveys. For the purpose of clustering analyses, such a population of galaxies acts as a probe of the underlying matter density field and allow efficient sampling of a large volume. The halo model is extremely useful for interpretation of clustering statistics. On small scales, clustering signal is described as being within the one-halo regime in which the clustering statistic is regarded as describing the clustering between objects in the same

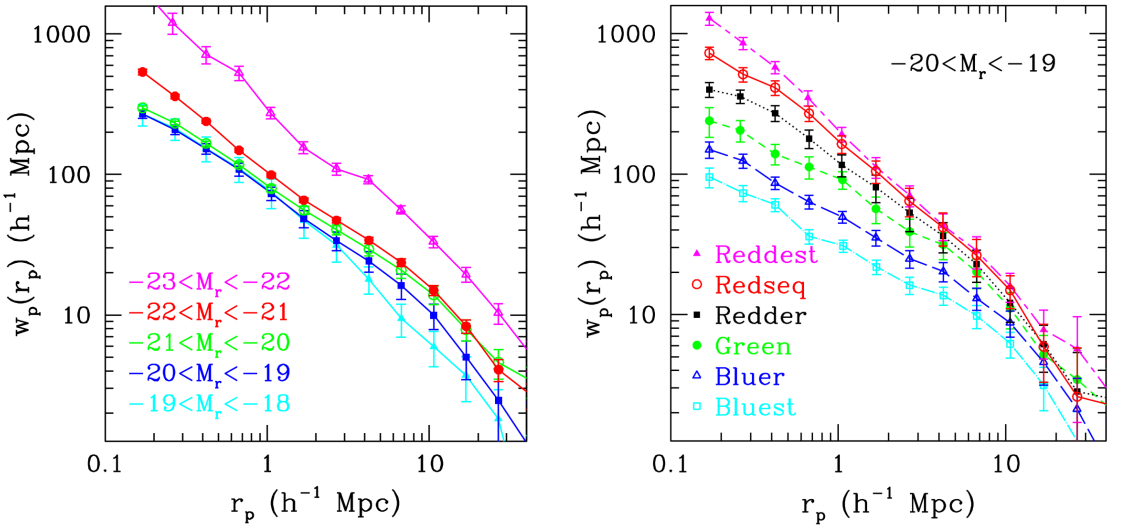

Figure 1.3: Projected correlation functions for luminosity (left) and colour (right) selected galaxy samples from the completed SDSS. Image credit: Zehavi et al. (2011).

halo. On larger scales, in the two-halo regime, clustering between objects in different halos is considered.

The large-scale distribution of matter is strongly cosmology-dependent and so has played a vital role in discriminating between cosmological models. Early work on the clustering of galaxies in the Automated Plate Measuring (APM) and QCD/QDOT surveys showed that the distribution of galaxies in the local Universe is inconsistent with what was then the standard CDM model, with $\Omega_{\rm M} = 1$ (Efstathiou et al., 1990; Saunders et al., 1991). Instead, the clustering on largest scales probed by these surveys was found to favour a cosmology with $\Omega_{\rm M} < 1$, providing the first evidence to point towards the currently accepted ΛCDM model.

The clustering results from many more galaxies surveys have since helped bolster support for the ΛCDM cosmology. Two notable examples are the power spectrum measurements of galaxies in the 2dFGRS and the SDSS (Tegmark et al., 2004; Cole et al., 2005; Tegmark et al., 2006). Clustering measures based upon data from these two surveys were amongst the first to identify the 'wiggles' in the galaxy power spectrum resulting from the baryon acoustic oscillations (BAO) in the early Universe. Just like in the CMB, the position of these acoustic peaks is a key prediction of ΛCDM. In the early Universe, the collapse of dark matter perturbations leads to the attempted collapse of perturbations in the baryon matter density field. However, prior to the epoch of recombination the baryons and radiation are coupled and so the attempted collapse of the baryon per-

turbations is impeded by the pressure support of the radiation. The oscillation in the radius of the perturbation therefore excites sound waves that propagate radially outwards from the perturbation at the local sound speed in the relativistic plasma. At the epoch of recombination, equal to $z \sim 1000$, the recombination of ions and electrons to a neutral gas leads to a decoupling of the radiation from the baryons, which abruptly reduces the sound speed and ends the propagation of the sound waves. Perturbations in the baryon matter density field are now free to collapse and the sound waves, the different modes of which will have completed a different number of oscillations, give rise to a series of harmonic maxima and minima in the baryon matter density power spectrum.

In the galaxy correlation function, these oscillations translate into a characteristic peak at a co-moving scale of ~ 150Mpc, equal to the sound horizon at that particular epoch (Eisenstein & Hu, 1998). Measurement of the position and amplitude of the peak can therefore provide conclusive evidence for the growth of perturbations (by linear perturbation theory) since $z \sim 1000$ and the existence of non-baryonic dark matter. Additionally, measurement of the BAO peak provides a reference scale of known length at any epoch, thus allowing the BAO peak to be used as a cosmic standard ruler. Measurement of the scale of the BAO peak at various epochs can therefore provide a robust distance-redshift measurement and an estimate of the angular diameter scale at that particular epoch (Blake & Glazebrook, 2003). By comparing these measurements to the angular scale in the CMB, the cosmic expansion history can be determined. Measuring the position of the BAO peak and using this information to constrain the dark energy equation of state, $w_{\rm DE}$, is therefore one of the major driving forces behind the next generation of galaxy surveys, including the Baryon Oscillation Spectroscopic Survey (BOSS Schlegel et al., 2009; Eisenstein et al., 2011; Sánchez et al., 2012), the WIGGLEz survey (Drinkwater et al., 2010), the Dark Energy Survey (DES Mohr et al., 2012; Abdalla et al., 2012) and the EUCLID survey (Laureijs et al., 2011). Although Eisenstein et al. (2005) were able to measure the BAO peak in a sample of low redshift ($z \simeq 0.35$) luminous red galaxies (LRGs) from the spectroscopic sample of the SDSS and determine $w_{\rm DE} = -1$ to ~ 10 per cent accuracy, their BAO detection was rather noisy due to the size of the SDSS sample and the large scales that need to be probed. At high redshifts, the BAO peak will not have been broadened as much by non-linear structure formation and so should allow more precise measurements of its position. The challenge for next generation surveys, therefore, is to measure the BAO peak in the correlation function of high redshift galaxies ($z \sim 2$) and place greater constraints on this measurement. As such, the measurement

of the observed clustering of galaxies continues to be of vital importance for constraining fundamental cosmological parameters.

Our current understanding is that, on smaller scales, the properties of galaxies are correlated with mass of the dark matter halo that is hosting them. For instance, LRGs are thought to be hosted by very massive dark matter structures (e.g. Wake et al., 2008). In this respect, we therefore describe a population of galaxies as being a biased tracer of the mass. Work by Norberg et al. (2001, 2002); Zehavi et al. (2002); Coil et al. (2004b); Zehavi et al. (2005); Wang et al. (2008); Zehavi et al. (2011), has clearly demonstrated that more luminous galaxies, and galaxies with redder colours, are more strongly clustered than less luminous, blue galaxies. Figure 1.3 shows the projected two-point correlation for various luminosity and colour selected galaxy samples in the completed SDSS (DR7 Abazajian et al., 2009). For luminosity selected samples, the clustering amplitude increases on all scales with increasing luminosity, with the clustering of $L > L^{\star}$ ($M_{\rm r}^{\star} = -20.44$) galaxies showing a more rapid increase than that of $L < L^{\star}$ galaxies. Similarly, the clustering amplitude of redder galaxies increases over that of bluer galaxies, with red galaxies displaying a steeper correlation function at small scales ($r_{\rm p} \sim 1 - 3h^{-1}$Mpc). The emerging picture from such analyses is one in which red, early-type galaxies can be found predominantly in galaxy clusters in high mass halos, whilst blue, late-type galaxies typically occupy more average locations, sometimes referred to as the field. Clearly, therefore, developing a working theory of galaxy formation, relating the formation and subsequent evolution of galaxies to the dark matter halos in which they reside, is of vital importance.

1.3.3 Redshift measurement and selection techniques

The measurement of the redshift of a galaxy can be made in one of two ways: either spectroscopically, by identification of a known feature in the spectrum of the galaxy, or photometrically, by measuring the flux in multiple wavebands and then, for instance, comparing this to the photometry of a set of galaxies with known redshifts.

The photometric approach works by using the flux in the multiple wavebands to provide a crude sampling of the spectral energy distribution (SED) of a galaxy. Broad features in the SED of a galaxy, such as the 4000Å break, mean that the colours of a population of galaxies, at a particular redshift, will occupy a small region in the multi-dimensional colour space. The redshift-dependence of this region means that an estimate of the redshift of a galaxy can be made by simply measuring its broad-band colours (e.g.

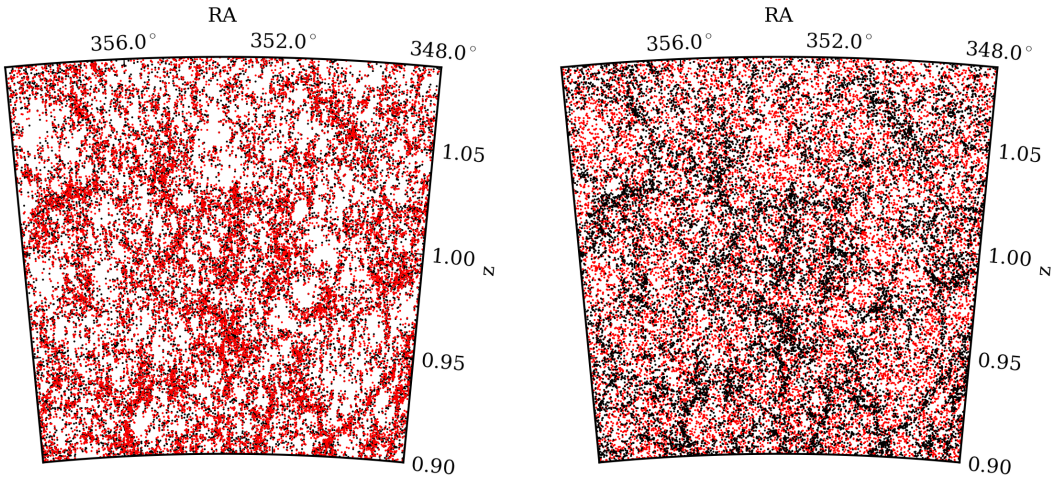

Figure 1.4: Comparison of the spatial distribution of galaxies in a synthetic Hα-selected galaxy survey using spectroscopic redshifts with an assumed error of $\sigma_z = 0.001(1 + z)$ (left) and using photometric redshifts with an assumed error of $\sigma_z = 0.05(1 + z)$ (right). In both panels, the black points show the true redshifts of galaxies and the red points show the effect of introducing the corresponding error into the redshift measurement.

Loh & Spillar, 1986; Connolly et al., 1995). Several techniques have been developed in an attempt to calibrate and optimise this procedure, including the use of SED template fitting (e.g. Bolzonella et al., 2000; Arnouts et al., 2002; Babbedge et al., 2004; Brammer et al., 2008; Assef et al., 2008), Bayesian priors (e.g. Benítez, 2000; Feldmann et al., 2006) and artificial neural networks (e.g. Firth et al., 2003; Collister & Lahav, 2004). Abdalla et al. (2011) compare several different codes for estimating photometric redshifts using the above techniques. Abdalla et al. find that, although the performance of each code depends upon the figure of merit used to assess it, the codes that make use of a training set of galaxies (the latter two techniques above) perform better at intermediate redshifts where complete training sets are available.

Both spectroscopic and photometric redshift measurements have their advantages and disadvantages. The main advantage of spectroscopic redshift estimates is their greater accuracy, with typical errors of less than ~ 0.1 per cent compared to errors of $3 - 10$ per cent on redshift estimates made using broadband photometry (e.g. Wolf et al., 2001; Csabai et al., 2003; Wolf et al., 2003; Collister et al., 2007; Rowan-Robinson et al., 2008; Budavári, 2009). Note also, that photometric redshift errors typically vary with galaxy type and the coverage in the rest-frame of the filter set used. The main disadvantage of photometric redshifts, therefore, is their reduced reliability. As Fig. 1.4 demonstrates,

the greater uncertainty in photometric redshift estimates leads to our view of the large-scale structure of galaxies being washed out, which increases the interloper fraction of galaxies at any particular redshift and the uncertainty in the clustering measurement at this redshift. The main disadvantage of spectroscopic redshift estimates, however, is that obtaining a spectrum of a galaxy is much more difficult and costly than simply measuring its broad band flux, especially for faint galaxies at high redshifts. Thus, many of the next generation of galaxy surveys that aim to measure the positions of millions of galaxies at high redshift ($z \lesssim 2$) are forced to use photometric redshifts.

An alternative approach to select populations of galaxies is to select galaxies based upon their broad band colours. Like photometric redshifts, these colour selection techniques rely upon the fact that broad features in the SEDs of galaxies will typically restrict the galaxies to a particular region of colour space. Such techniques are commonly used to identify populations of galaxies at high redshift ($1 \lesssim z \lesssim 3$), which was difficult to probe prior to the development of near-infrared spectrographs. Colour selection techniques provide a relatively inexpensive method for selecting large numbers of these high redshift galaxies.

A notable example of a high-redshift selection technique is the Lyman-break dropout technique (Steidel et al., 1996, 2003, 2004). The large ionisation cross-section of neutral hydrogen leads to a prominent break in the rest-frame SEDs of highly star-forming galaxies at the Lyman limit of 912Å. For galaxies at $z \sim 3$, this break is shifted to a wavelength of $\lambda \sim 4000$Å, which falls between the U and B-bands. The drop out technique is therefore used to identify highly star-forming galaxies at $z \sim 3$ (known as *Lyman Break Galaxies*, LBGs) by searching for galaxies that are detected in the B-band but are undetected in the U-band. Other selection techniques have been devised that are capable of identifying faint, extremely red galaxies at $z \sim 1$ (Elston et al., 1988; McCarthy, 2004) and distant red galaxies, thought to be amongst the most massive galaxies in the $z \sim 2$ population (Franx et al., 2003). An additional example, which we will discuss further in Chapter 4, is the BzK technique, which uses the B, z and K-bands to identify both star-forming and passively evolving galaxies at $1.5 < z < 2.5$ (Daddi et al., 2004a).

Increasingly, synthetic galaxy catalogues are playing a key role in the analysis of survey data. Such catalogues can be constructed empirically or from theoretical models, as we do in this thesis.

1.4 Numerical techniques

This thesis focuses on the construction of lightcone mock catalogues using simulations of structure and galaxy formation. We now outline the main techniques employed in this area.

Our understanding of the development of dark matter structures, known as *halos*, from the initial growth of primordial perturbations to their eventual collapse under gravity is well understood and can be reproduced in N-body simulations with reasonable confidence. Cosmological N-body simulations, such as the *Millennium Simulation* (Springel, 2005) and the *Millennium-XXL Simulation* (Angulo et al., 2012, see Fig. 1.5), have enabled spectacular visualisations of the hierarchical development of the filamentary large-scale structure of the Universe that has been revealed through galaxy surveys such as the 2dFGRS and SDSS. In these simulations the most over-dense structures, host rich clusters of galaxies. Smaller over-densities, which undergo collapse at later times, form less massive halos that may go on to host small groups of galaxies or individual galaxies. The less massive halos accumulate in the filaments and sheets. The gravitational attraction of the most massive halos means that the less massive halos fall into the inter-connecting filaments and are ultimately accreted onto the more massive halos at the nodes. As a result of their gravitational attraction, the nodes and filaments become increasingly over-dense, thus causing the voids in-between to become increasingly under-dense.

The main issues facing cosmological simulations now, however, are the problem of resolution, due to limitations in computing resources, and the treatment of the processes that govern the evolution of baryons.

1.4.1 Resolution vs. volume

Limitations in resolution mean that one must have an appreciation of the scales that one wishes to probe with the simulation beforehand. Placing a fixed number of particles into a larger simulation box, means that each particle must carry a greater mass. As a result, using a box with larger dimensions means that one is restricted to working with more massive dark matter structures, such as, for example, massive dark matter halos that host clusters of galaxies, rather than the less massive halos that host individual galaxies. Most simulations capable of resolving halos of mass $\sim 10^{12}h^{-1}\mathrm{M}_{\odot}$, have box sizes of approximately 100-500 h^{-1}Mpc on a side. The problem is driven however, by the need to simulate larger cosmological volumes, comparable to the volumes that will be mapped

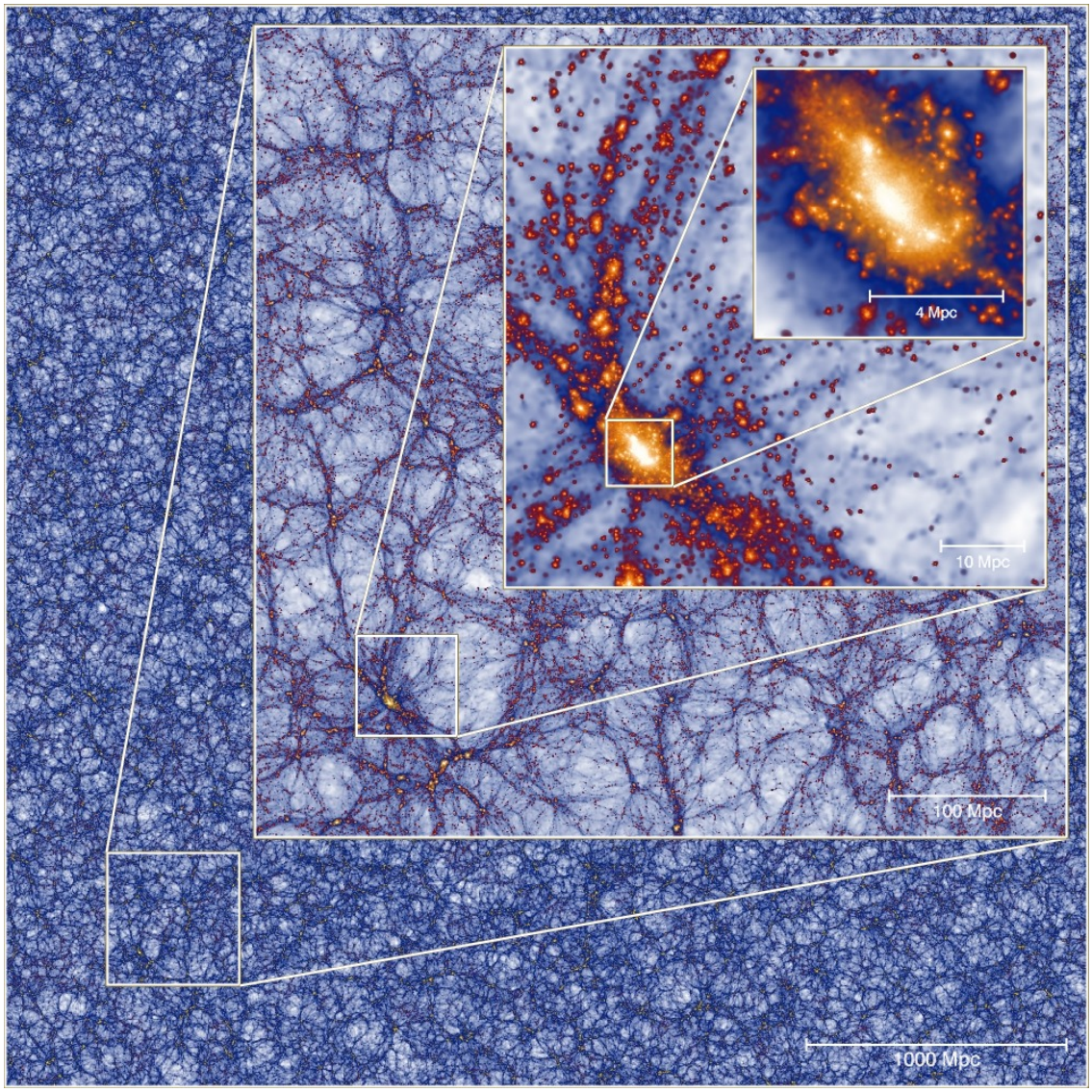

Figure 1.5: Zoomed images demonstrating the large range of scales probed by the *Millennium-XXL Simulation* (Angulo et al., 2012). Image credit: R. Angulo, Virgo Consortium.

by up-and-coming galaxy surveys. For instance, the largest mode whose evolution can be trusted in a simulation corresponds to roughly one tenth of the dimension of the simulation box. When examining the clustering of galaxies in a simulation, one is limited to examining scales no larger than this. However, the BAO peak discussed above appears in the clustering signal of galaxies at approximately 150 Mpc, meaning that to realistically recover the signal at these scales, one will need a simulation with box size of at least 1-$2h^{-1}$Gpc. Typically, therefore, a compromise has to be made between the volume one wishes to simulate and the mass resolution at which one wishes to sample this volume.

1.4.2 Including the baryons

For simulating the baryon content of the Universe, two complementary approaches have emerged: direct hydrodynamical simulations and semi-analytic modelling.

In hydrodynamical simulations, which attempt to directly solve numerically the hydrodynamic equations governing the baryons, typically one of two approaches is adopted. The first, known as *smooth particle hydrodynamics* (SPH), is a Lagrangian approach that uses particles to trace the baryons in a similar way to that in which dark matter is traced in N-body simulations (though, unlike N-body simulations, SPH simulations differ in that the particle positions are smoothed). The complication now, however, is that, unlike dark matter, the baryons are not dissipationless and so will be affected by changes in temperature and pressure induced by radiative processes. The second approach, known as *adaptive mesh refinement* (AMR), instead sets up an adaptive grid and uses an Eulerian approach to trace the hydrodynamical properties of baryon flux in and out of the various grid cells. Just like dark matter simulations, hydrodynamical simulations suffer from limited resolution, with many of the physical processes affecting the baryons, such as star formation, acting on scales below the resolution limit of the simulation. This has led to many of the processes being described as "sub-grid physics", which are decoupled from the hydrodynamical processes.

The alternative method, pioneered by the early work of White & Rees (1978); White & Frenk (1991); Cole et al. (1994), is that of *semi-analytic modelling*, in which the evolution of the baryon content of dark matter halos is followed using simple differential equations to describe the physical processes that affect the baryons. In many cases, our current understanding of the underlying physics is crude, leading to many of the equations containing parameters that account for our ignorance. Unfortunately, our limited knowledge of the physics means that the values of these parameters cannot be derived from first prin-

ciples. Instead, the values of the parameters are physically motivated wherever possible and are constrained by requiring that the model be able to reproduce a given subset of observables of the galaxy population. Typically, these observables correspond to various statistics of local galaxies, such as the galaxy luminosity function or the Tully-Fisher relation. Previously, semi-analytic models failed to simultaneously reproduce the faint-end slope and bright-end fall-off of the luminosity function. This has now been resolved by invoking supernova and AGN-driven feedback mechanisms (Benson et al., 2003; Croton et al., 2006; Bower et al., 2006). Bower et al. additionally showed that inclusion of AGN feedback enables models to predict a galaxy colour-magnitude relation that is bimodal, in agreement with that observed in the SDSS. To date, however, models still struggle to simultaneously fit the luminosity function and the zero-point of the Tully-Fisher relation, typically predicting disk rotation velocities that are too high[5].

One of the main advantages that semi-analytic models have over hydrodynamical simulations is that they are comparatively inexpensive computationally. This allows the models to be run much faster than hydrodynamical simulations, over larger cosmological volumes. Additionally, unlike direct simulations, the physics in semi-analytic models can be readily adjusted to allow a rapid exploration of the parameter space. Although semi-analytic models are criticised for carrying a large number of free parameters, the predictions of the models have been shown to be sensitive to only a small number of these parameters, implying the robustness of the models. In addition, the modular construction of semi-analytic models means that new physics and numerical techniques can be readily incorporated. Overall the difference between semi-analytic models and hydrodynamical simulations is becoming less clear cut, with hydrodynamical simulations being forced to adopt similar parameterised prescriptions, whilst semi-analytic models are beginning to incorporate more general algorithms that solve for physical processes numerically, (e.g. the star-formation modelling of Lagos et al., 2011b). The inclusion of these algorithms may remove the need for some parameters whilst simultaneously narrowing the available parameter space for the remaining parameters. The fact that incorporating such changes has little effect on the model predictions bolsters support for the reliability of semi-analytic models.

[5]Currently, the problem of simultaneously matching the luminosity function and the zero-point of the Tully Fisher relation appears to stem from the fact that dark matter halos have density profiles that are too centrally concentrated. It is not yet clear whether this is a problem for the CDM cosmology, or whether current models for disk formation are incorrect.

1.4.3 Mock galaxy catalogues

One of the main problems when comparing observational results to simulated predictions is that simulations and observations naturally work in very different frameworks; whilst simulations work with the idealised case of a complete, volume-limited sample of galaxies positioned in real-space, observational galaxy samples are primarily flux-limited samples in redshift-space which, due to selection effects, may be only partially complete. This must be appreciated when comparing observational results to simulation predictions such that any biases or systematics introduced by either dataset are properly considered. Ideally, a comparison should be done by converting one dataset into the framework of the other; typically by translating the simulation predictions into the framework of the observations. This is preferable for the obvious reason that it is easier to introduce observational selection effects into the simulation than it is to attempt to remove them from the observations, i.e. "it is easier to convolve than to deconvolve". Additionally, much more information about galaxy properties is available from a simulation than from the observations to make translating the simulation the more viable option. This necessity to translate the simulation into the observational framework gives rise to the concept of *mock galaxy catalogues*, which form the central focus of this work.

When working with an observational galaxy catalogue, an estimator designed to recover a statistic, such as the luminosity function or correlation function, will have to compensate for a variety of effects such as non-uniform coverage of the sky and a selection function that varies strongly with radial distance from the observer. The primary advantage of a mock catalogue is that, by construction, we already know the 'true' answer for the statistic without these effects. By comparing a measurement extracted from a synthetic mock catalogue with the ideal result (i.e. the statistic measured using a complete sample of galaxies from the original simulation cube), one can adjust and tune the performance of the estimator to reduce any systematic effects. A prime example is that of algorithms designed to find groups of galaxies, the calibration of which requires foreknowledge of the underlying dark matter halo distribution in order to test how faithfully the algorithm can recover these structures when working in redshift space (Eke et al., 2004a; Robotham et al., 2011; Murphy et al., 2012). Additionally, mock catalogues can be used to forecast the scientific return of future galaxy surveys (Cole et al., 1998; Cai et al., 2009; Orsi et al., 2010). Therefore they can help shape the design of a survey by assessing the level, and quality, of the statistics recoverable with any particular configu-

ration. Finally, mock catalogues allow us to cast the predictions of theoretical models of galaxy formation in a form that can be directly compared against observables.

The size of modern galaxy surveys has heralded the start of an era of precision cosmography wherein we can measure statistics, such as the galaxy luminosity function, with random errors that are smaller than the systematic errors. To continue to make progress it is essential that we improve our understanding of how the estimation of such statistics is affected by the construction of a galaxy survey and the selection criteria applied. Mock galaxy catalogues, which mimic the selection effects in real surveys, have therefore emerged as an essential tool with which to achieve this aim, and play a central role in the analysis and exploitation of galaxy surveys.

1.5 Motivation & outline

Our motivation is to show that mock catalogues are extremely useful for interpretation of observational data and relating observations to the predictions of galaxy formation models. The main theme of this thesis is the construction of a new code to build mock catalogues from semi-analytic galaxy formation models combined with large volume N-body simulations. Mock catalogues are accepted as essential tools in the set-up and post-processing analysis of galaxy surveys.

The outline for the rest of this thesis is as follows:

- In Chapter 2 we introduce the `GALFORM` semi-analytical model, the galaxy formation model that we will use throughout this thesis.

- In Chapter 3 we discuss how we use the output from `GALFORM` to build lightcone galaxy mock catalogues.

- In Chapter 4 we apply an example lightcone to investigate the effectiveness of the BzK colour selection technique.

- In Chapter 5 we use a simplified mock catalogue to examine the calibration of a friends-of-friends group finder for application to the Six-degree Field Galaxy Survey (6dFGS).

- In Chapter 6 we use a lightcone to look at trends in the angular correlation function of galaxies in future photometric redshift surveys.

- In Chapter 7 we outline our final conclusions and summarise the work that we have presented.

Chapter 2

The GALFORM *semi-analytical model of galaxy formation*

The Durham semi-analytical galaxy formation model, GALFORM, originally developed by Cole et al. (2000), models the star formation and merger history of a representative population of galaxies and makes predictions for many galaxy properties including luminosities over a substantial wavelength range extending from the far-UV through to the sub-millimetre (Baugh et al., 2005; Lacey et al., 2008, 2010; Lagos et al., 2011a; Fanidakis et al., 2011; Lagos et al., 2012). In this chapter we will describe the various details of the GALFORM model, that we use throughout this thesis.

The GALFORM model populates a distribution of dark matter halos with galaxies by using a set of coupled differential equations to determine how, over a given time-step, the physics that regulates the sizes of and exchange of material between the various baryonic components of galaxies. GALFORM models the main physical processes governing the formation and evolution of galaxies: (i) the collapse and merging of dark matter (DM) halos, (ii) the shock-heating and radiative cooling of gas inside DM halos, leading to the formation of galactic disks, (iii) quiescent star formation in galactic disks, (iv) growth of supermassive black holes, (v) feedback as a result of supernovae, active galactic nuclei and photo-ionisation of the inter-galactic medium, (vi) chemical enrichment of stars and gas, (vii) dynamical friction driven mergers of galaxies within DM halos and (viii) a treatment of self-gravitating galactic disks that have become unstable. In the GALFORM model it is these last two processes (vii and viii) that lead to the formation of spheroids and triggering of starburst events. Each of these processes will be discussed in the following sections.

2.1 Dark matter halos

Crucial to any theory of galaxy formation is an understanding of the formation of a population of dark matter halos. Compared to the baryonic processes governing the formation and subsequent evolution of galaxies, the formation of dark matter halos is understood fairly well and can be modelled relatively easily.

The merger histories of a population of dark matter halos can currently be generated through one of two methods: using a Monte-Carlo approach to sample the predicted mass function of halos at different epochs, or by directly following the hierarchical build-up of structure in a N-body simulation.

2.1.1 Monte-Carlo merger trees

Generating a population of dark matter halos using the Monte-Carlo approach involves statistically sampling the mass function of halos at various epochs, as predicted by extended Press-Schechter theory (Bond et al., 1991; Bower, 1991), and deriving a merger rate for halos between these epochs (e.g. Lacey & Cole, 1993).

Press & Schechter (1974) were the first to analytically predict the abundance of gravitationally bound halo-like structures. By applying a spherical top-hat window function to a Gaussian matter density field, Press & Schechter were able to predict the fraction of mass in collapsed structures as the fraction of peaks within the field with an over-density[1], δ, greater than or equal to some critical value, $\delta_{\rm c}$. The Press-Schechter mass function for the number of gravitationally bound structures, with mass in the range M to $M + {\rm d}M$ is,

$$\frac{{\rm d}n(M,t)}{{\rm d}M} = \sqrt{\frac{2}{\pi}} \frac{\rho_0}{M^2} \frac{\delta_{\rm c}(t)}{\sigma(M)} \exp\left(-\frac{\delta_{\rm c}^2(t)}{2\sigma^2(M)}\right) \left|\frac{{\rm d}\ln\sigma(M)}{{\rm d}\ln M}\right|, \tag{2.1}$$

where ρ_0 is the mean density of the Universe, $\delta_{\rm c}(t)$ is the critical linear over-density necessary for a structure to collapse at a time t (e.g. for a universe with $\Omega(z=0)=1$ today, $\delta_{\rm c}(z=0) \simeq 1.686$) and

$$\sigma^2(M) = \frac{1}{2\pi^2} \int_0^\infty P(k) \tilde{W}^2(k) k^2 {\rm d}k \tag{2.2}$$

is the mass variance of the smoothed matter density field with power spectrum $P(k)$. In Eq. (2.2), $\tilde{W}(k)$ is the Fourier transform of the real-space, spherical top hat window function $W(R)$. The Press-Schechter formalism however, neglects under-dense regions

[1]The linear theory density contrast of a matter field is defined as $\delta = \rho(\mathbf{x},t)/\rho_0(t) - 1$, where $\rho(\mathbf{x},t)$ is the density field at position $\mathbf{x}$ and time t and $\rho_0(t)$ is the mean density of the matter field.

that may be enclosed by larger over-dense regions and so have a non-zero chance of collapsing. As such, only half of the mass in the Universe is accounted for. This formed part of the *cloud-in-cloud* problem (e.g. Peacock & Heavens, 1990), which led to the development of extended Press-Schechter (EPS) theory (Bond et al., 1991; Bower, 1991). The conditional mass function given by the EPS theory,

$$f(M_1|M_2)\mathrm{d}\ln M_1 = \sqrt{\frac{2}{\pi}}\frac{\sigma_1^2(\delta_1-\delta_2)}{\left[\sigma_1^2-\sigma_2^2\right]^{3/2}}\exp\left[-\frac{1}{2}\frac{(\delta_1-\delta_2)^2}{\left(\sigma_1^2-\sigma_2^2\right)}\right]\left|\frac{\mathrm{d}\ln\sigma}{\mathrm{d}\ln M_1}\right|\mathrm{d}\ln M_1, \qquad (2.3)$$

provides the fraction, $f(M_1|M_2)$, of mass in gravitationally bound structures with mass M_2 at redshift z_2, that was contained in the progenitors of these structures at an earlier redshift, $z_1 > z_2$, with a mass M_1. The linear theory critical densities at epochs z_1 and z_2 are given by δ_1 and δ_2 respectively, and σ_1 and σ_2 represent Eq. (2.2) evaluated for the masses M_1 and M_2 respectively (at the epochs z_1 and z_2 respectively).

By taking the limit of Eq. (2.3) as $z_1 \to z_2$, one can estimate the mean number of objects, each of mass M_1, that will merge together over a time interval $\mathrm{d}z_1$ to form an object of mass $M_2 > M_1$, according to

$$\frac{\mathrm{d}N}{\mathrm{d}M_1} = \frac{\mathrm{d}f(M_1|M_2)}{\mathrm{d}z_1}\frac{M_2}{M_1^2}\mathrm{d}z_1. \qquad (2.4)$$

The Monte-Carlo algorithm employed by GALFORM follows the basic steps:

1. Specify a mass, M_2, and redshift, z, for the final halo in the Monte-Carlo merger tree.

2. Specify a mass resolution limit, $M_{\rm res}$, for the halos. Then define the mean number of progenitors of the halo, with masses in the interval[2] $M_{\rm res} < M_1 < M_2/2$, as,

$$P = \int_{M_{\rm res}}^{M_2/2}\frac{\mathrm{d}N}{\mathrm{d}M_1}\mathrm{d}M_1, \qquad (2.5)$$

   and the fraction of the mass of the final halo that was previously contained in progenitors below the mass resolution limit as,

$$F = \int_0^{M_{\rm res}}\frac{\mathrm{d}N}{\mathrm{d}M_1}\frac{M_1}{M_2}\mathrm{d}M_1. \qquad (2.6)$$

3. Select a redshift interval, $\mathrm{d}z$, so that the halo is unlikely to have more than two progenitors at the earlier epoch, $z + \mathrm{d}z$, and $P \ll 1$.

[2]Note that this method only works if progenitors are drawn with a mass $M_{\rm res} < M_1 < M_2/2$.

4. Generate a uniform random number $R = [0, 1]$. If $R > P$, the halo is not split into two progenitors but simply has its mass reduced by a factor of $1 - F$. However, if $R \leqslant P$ then a random value mass M_1 (in the range $M_{\rm res} < M_1 < M_2/2$) is generated and the halo is split into two progenitors of mass M_1 and $M_2(1 - F) - M_1$.

5. Repeat the above steps 1-4 for each progenitor in successive redshift steps until the full halo tree is constructed.

Parkinson et al. (2008) introduced a perturbing function into Eq. (2.4) so that it is made consistent with EPS theory in the limit that $M_1 \rightarrow (M_2 - M_1)$.

2.1.2 N-body halos

Dark matter particles are thought to interact with other forms of matter through gravity alone. As a result, dark matter structures can be modelled as dissipationless particles in a N-body simulation.

Compared to the Monte-Carlo approach to building merger histories, the use of N-body simulations has several advantages, thanks principally to the provision of the spatial and velocity information for halos. Firstly, with this information the clustering statistics for the halo population can be successfully extracted. Additionally, on smaller scales, one is able to make predictions for the positions of galaxies within their host halos. Secondly, one of the fundamental assumptions underpinning the Monte-Carlo approach is that the formation history of a dark matter halo depends simply on its mass and is independent of the 'environment' in which the halo is found. However, studies of halos in high-resolution N-body simulations have suggested that the clustering amplitude of Milky Way sized halos is dependent upon the redshift at which the halo formed (e.g. Gao et al., 2005). Tracing the formation histories of a population of halos in a simulation therefore means that other factors affecting their formation history (such as 'environment') are properly incorporated. One of the main disadvantages, however, of the N-body approach is the limited dynamical range of halos that can be simulated within a large cosmological volume. This is due to limitations in the computational resources that are available. Furthermore, as we shall see in Section 2.2.2, the identification and tracking of substructures within the simulation is not without its problems.

We will discuss the construction of a typical cosmological, N-body simulation by considering, as an example, the *Millennium Simulation*, a 2160^3 particle N-body simulation of the ΛCDM cosmology that was carried out by the Virgo Consortium (Springel et al.,

2005). For future reference, we note that this is the simulation that we will use in conjunction with our chosen semi-analytical model.

2.2 The Millennium Simulation

The *Millennium Simulation* follows the hierarchical growth of cold dark matter structures from redshift $z = 127$ through to the present day in a cubic volume of $500h^{-1}$Mpc on a side. The cosmological parameters adopted in the simulation are: a baryon matter density $\Omega_{\rm b} = 0.045$, a total matter density $\Omega_{\rm m} = \Omega_{\rm b} + \Omega_{\rm CDM} = 0.25$, a dark energy density $\Omega_\Lambda = 0.75$, a Hubble constant $H_0 = 100h\,{\rm km\,s^{-1}Mpc^{-1}}$ where $h = 0.73$, a primordial scalar spectral index $n_{\rm s} = 1$ and a fluctuation amplitude $\sigma_8 = 0.9$. These parameters were chosen to match the cosmological parameters estimated from the first year results from the Wilkinson Microwave Anisotropy Probe (WMAP1, Spergel et al., 2003).

Halo merger trees are constructed using particle and halo data stored at 64 fixed epoch snapshots that are spaced approximately logarithmically in expansion factor. The Millennium trees have a temporal resolution of approximately 0.26 Gyr at the present day, with approximate resolutions of $0.38, 0.35$ and $0.26\,{\rm Gyr}$ at redshifts $z = 0.5, 1$ and 2 respectively. Halos in the simulation are resolved with a minimum of 20 particles, corresponding to a halo resolution of $M_{\rm halo,lim} = 1.72 \times 10^{10} h^{-1}{\rm M}_\odot$, which is significantly smaller than that expected for the Milky Way's dark matter halo.

2.2.1 Constructing a N-body simulation

The linear theory matter power spectrum for the WMAP1 cosmology was computed using the code CMBFAST (Seljak & Zaldarriaga, 1996). To construct the initial unperturbed density field, the 2160^3 particles were placed into the simulation volume with an initially Poisson particle distribution, which was subsequently evolved with the sign of gravity reversed to create a *glass-like* particle distribution, with negligible residual forces (Baugh et al., 1995; Springel et al., 2005).

The dynamical evolution of the particles under the action of gravity is described by the Hamiltonian,

$$H = \sum_i \frac{\mathbf{p}_i^2}{2m_i a(t)^2} + \frac{1}{2}\sum_{ij} \frac{m_i m_j \phi(\mathbf{x}_i - \mathbf{x}_j)}{a(t)} \tag{2.7}$$

where $H = H(\mathbf{p}_1, ..., \mathbf{p}_N, \mathbf{x}_1, ..., \mathbf{x}_N, t)$, $\mathbf{x}_i$ are the co-moving coordinate vectors of the particles, $\mathbf{p}_i = a(t)^2 m_i \dot{\mathbf{x}}^2$ are the canonical momenta of the particles and $\phi(\mathbf{x})$ is the

gravitational interaction potential. Time evolution is built in through the dependence upon the expansion factor, $a(t)$.

The dynamics of the system can be described by solving the Poisson equation,

$$\nabla^2\phi(\mathbf{x}) = 4\pi G\left[\rho(\mathbf{x}) - \bar{\rho}\right]. \tag{2.8}$$

The gravitational forces on the particles are calculated using a variant of the *TreePM* method (Barnes & Hut, 1986), which uses a hierarchical multipole expansion 'tree' algorithm to compute the short-range gravitational forces and a particle mesh (PM) to calculate the long-range forces. This involves placing the particles onto a density grid and then solving Eq. (2.8) using a *Fast-Fourier Transform* (FFT), i.e. solving $k^2\tilde{\phi}_k = 4\pi G\hat{\delta}_k$, where $\tilde{\phi}_k$ is the Fourier transform of the interaction potential and $\tilde{\delta}_k$ is the Fourier transform of the density distribution, $\delta(\mathbf{x})$. The positions of particles are updated by solving the Poisson equation for the gravitational potential at each timestep using the updated density field. The timesteps for individual particles are determined using a symplectic leapfrog scheme, with the short-range and long-range forces solved at interleaved time points (e.g. Springel et al., 2005). Knowing the interaction potential allows one to determine the gravitational forces acting on each particle. Summation of these forces allows the particle positions and velocities to be updated. In Eq. (2.8), the mean density, $\bar{\rho}$, is subtracted meaning that the solution to the Poisson equation relates to the peculiar potential and one can examine fluctuations about the mean density. The density contrast of the particles, $\delta_\varepsilon(\mathbf{x})$, as function of their positions, $\mathbf{x}$, is described over a finite scale equivalent to a gravitational softening length, ε. This softening length prevents the gravitational potential from diverging between close pairs of particles and ensures that each particle sees a 'smooth' background density. For the softening length chosen for the *Millennium Simulation*, the Newtonian gravitational potential of a point mass, m, on small scales is $-Gm/\varepsilon$, equivalent to a Plummer sphere with radius ε. For the *Millennium Simulation*, $\varepsilon = 5\,h^{-1}$kpc (Springel et al., 2005).

2.2.2 Identifying halos and constructing merger trees

To construct the halo merger trees one must first identify groups of dark matter particles in each of the simulation snapshots. This is done using the *Friends-Of-Friends* algorithm (FOF, Davis et al., 1985). The *Millennium Simulation* was carried out with a specially modified version of the GADGET2 code (Springel, 2005) with a built-in FOF group-finder, allowing FOF groups to be identified on the fly. The FOF algorithm identifies groups

of objects by searching within a spherical volume around each object. The volume is described using a *linking length*, l, which is defined as,

$$l = \frac{b}{n^{1/3}}, \tag{2.9}$$

where b is the *linking parameter*, usually given a value equal to a fraction of the mean inter-object separation, and n is the mean number density of objects of interest.

A value of $b \sim 0.2$ is often chosen as this has been shown to successfully identify structures with a density contrast of $\Delta_{\rm c} \sim 200$. For the spherical collapse model, this density contrast corresponds approximately to the radius at which the amount of material infalling onto the halo balances the amount of material escaping from the halo, which provides a suitable definition for isolating the material interior to the virialised halo from the surrounding material still undergoing infall (Cole & Lacey, 1996; Eke et al., 1996). Additionally, Jenkins et al. (2001) have shown that adopting a linking parameter of $b = 0.2$ (and identifying halos according to this fixed choice of overdensity, independent of the value of $\Omega_{\rm M}$), predicts a universal halo mass function whose shape appears to be independent of redshift, the shape of the matter power spectrum and the values of $\Omega_{\rm M}$ and Ω_{Λ}. However, since a standard FOF algorithm links particles based purely on their proximity to one another, the recovered FOF groups may not necessarily correspond to bound structures.

To identify the bound structures, each simulation snapshot is post-processed using a (sub-)halo finding algorithm. For the *Millennium Simulation*, the algorithm SUBFIND (Springel et al., 2001) was used to identify self-bound, locally over-dense sub-groups within the FOF groups. SUBFIND works by identifying local, self-bound maxima on top of the smoothed background density field of the dark matter. The background density field for each FOF group is usually taken to be the density field of the most massive sub-group. In practise, this is done by sorting the particles in the FOF group in terms of the density at the position of the particle. The density at the position of each particle is determined by kernel interpolation over the 10 nearest neighbours, starting with the particle with the highest density. As each particle is added back into the group, the algorithm searches for any two of the particle's 10 nearest neighbours that have a higher density. If none of the nearest neighbours have a higher density, then the particle becomes the centre of a new sub-group. If one or two neighbours have higher density and are attached to the same sub-group, then the particle is also associated with that sub-group. If two higher density neighbours are identified and are each attached to a different sub-group then the

particle is regarded as a saddle point between the background field and the maximum of the local field of another sub-group. Such maxima become candidates for possible sub-structures within the FOF group. The sub-group associated with each local maximum is then subjected to a gravitational unbinding procedure in order to reject those particles with a positive total energy that are not part of the true bound substructure. Sub-groups which contain a minimum of 20 bound particles become classified as identifiable sub-halos, while the remainder of the sub-groups are discarded. The application of the SUBFIND algorithm typically results in the bulk of the mass of a FOF group being assigned to one large sub-group which represents the background mass distribution of the halo. The remaining mass is usually split between smaller satellite sub-groups orbiting within the halo and unbound "fuzz" particles which are not associated with any sub-group.

However, it is not uncommon for the FOF algorithm to join together structures which might be better considered as separate halos for the purposes of semi-analytic galaxy formation. For example, nearby groups may be linked by tenuous "bridges" of particles or they may only temporarily be joined. The merger tree algorithm we use in this thesis is intended to deal with these cases and ensure that the resulting trees are strictly hierarchical, i.e. once two halos are deemed to have merged they should remain merged at all later times.

The first step in the construction of the merger trees is to identify a descendant for each sub-group at the next snapshot. The descendant of each sub-group is identified as the sub-group at the next snapshot that contains the largest number of the $N_{\rm link}$ most bound particles, where

$$N_{\rm link} = \max\left(f_{\rm trace} N_{\rm p}, N_{\rm linkmin}\right), \tag{2.10}$$

with $N_{\rm p} \geqslant 20$ and $f_{\rm trace}$ and $N_{\rm linkmin}$ are set to 0.1 and 10 respectively. Defining $N_{\rm link}$ in this way means that in well resolved cases we follow the most bound "core" of the sub-group, which is important for satellite sub-groups which may be tidally stripped of their outer parts. For the smallest groups with $N_{\rm p} \sim 20$, we have $N_{\rm link} = 10$ so that we are following at least 50% of the particles (and so preventing inaccurate assignment due to low number statistics).

The SUBFIND algorithm occasionally temporarily "loses" a sub-group between snapshots. For example, a sub-group may be identified at snapshot i, lost at one or more subsequent snapshots, and then identified again at snapshot $i+n$, where $n > 1$. This can happen if a small, isolated group briefly falls below the resolution limit of the simulation or if a satellite sub-group passes close to the centre of its host halo. In either case we

would like to identify the sub-group at snapshot $i+n$ as the descendant of the sub-group at snapshot i. Our approach to achieve this aim is as follows:

1. Identify sub-groups which may have been lost by SUBFIND.

2. Identify sub-groups which may have just been reacquired by SUBFIND.

3. Attempt to locate descendants of the sub-groups in (i) with the sub-groups in (ii).

Groups which are "lost" are identified by looking for groups which either have no immediate descendant or are not the most massive progenitor of their immediate descendant. Some of these groups will have been lost because they have genuinely been disrupted and absorbed into the parent halo, but some will reappear later. Groups which have just been reacquired are identified by looking for "orphan" sub-groups, i.e. groups with no immediate progenitors.

For each lost sub-group at snapshot i (where i is not the present day), we examine the orphan sub-groups at snapshot $i+2$, $i+3$, ... , $i+N_{step}$. An orphan sub-group is identified as the descendant of a lost sub-group if at least a fraction $f_{\rm link}$ of the $N_{\rm link}$ most bound particles from the lost group are in the orphan group and no orphan descendant can be found at earlier snapshots. We usually set $N_{\rm step}=5$ and $f_{\rm link}=0.5$.

If this procedure results in the identification of a descendant for a sub-group, then that descendant will be used in the subsequent stages of the construction of the merger trees. For all other sub-groups the descendant is taken to be the immediate descendant at the next snapshot. For the construction of merger trees, having a sub-group and its descendant separated by multiple snapshot outputs is not a problem. However, this is inconvenient for codes, such as GALFORM, which expect the descendant of a subhalo to always be found in the next snapshot. To avoid this, for those subhalos that are temporarily lost, interpolated sub-halos are inserted at each snapshot where the sub-halo is 'missing'. For very high resolution simulations this is a common occurrence. However, for simulations like the *Millennium Simulation* such interpolated sub-halos are rare.

Next, the sub-groups at each snapshot are organised into a hierarchy of halos, sub-halos, sub-sub-halos etc. For each sub-group in a FOF group we identify the least massive of any more massive "enclosing" sub-groups in the same FOF group. Sub-group A is said to enclose sub-group B if centre of B lies within twice the half mass radius of A. Any sub-group which is not enclosed by another is considered to be an independent halo. We also consider a sub-group to be an independent halo if it has retained at least 75 per cent

of the maximum mass it had ever had whilst being the most massive sub-group in its FOF group. This is because we expect a halo involved in a genuine merger with a more massive halo to be stripped of mass. In either case, if a sub-group is deemed to be an independent halo then any sub-groups it encloses are also assigned to that halo.

At this stage we have, for each snapshot, a population of halos, each of which consists of a grouping of SUBFIND sub-groups with pointers linking each sub-group with its descendant at the next snapshot. We choose the descendant of a halo to be the halo at the next snapshot which contains the descendant of the most massive sub-group in the halo. This defines the halo merger tree structure.

In general, the GALFORM model assumes that when a halo merges with another, more massive "host" halo, that its hot gas is stripped away so that no further gas can cool in the less massive halo. Since a halo can only be stripped of hot gas once, we wish to treat these objects as satellite sub-halos within their host halo for as long as they survive in the simulation, even if their orbit puts them outside the virial radius of their host halo at some later times. We therefore attempt to identify cases where halos fragment, and re-merge them.

In practice we implement this by looking for satellite sub-groups which split off from their host to become independent halos at the next snapshot. A sub-group will be re-merged if it satisfies all of the following conditions:

- The sub-group is the most massive progenitor of its descendant. This is taken to mean that the sub-group survives at the next snapshot.
- The sub-group is not the most massive sub-group in its halo. This indicates that it is a satellite sub-halo within a larger halo.
- The descendant of the sub-group is the most massive group in its halo.
- The descendant of the sub-group belongs to a halo other than the descendant of the halo containing the original sub-group. This indicates that the host halo has fragmented.

The last condition is necessary because a sub-group can sometimes become the most massive in its parent halo without any halo fragmentation occurring, especially if the halo consists of two sub-groups of similar mass. If these conditions are met, the halo containing the descendant of the satellite sub-group is merged with the descendant of the host halo.

Following this post-processing, we are left with, for the *Millennium Simulation*, approximately 20 million halo merger trees with, in total, approximately 1 billion nodes.

2.3 Halo properties

Knowledge of the internal structure of halos is necessary to determine the properties of the galaxies that these halos will host. The principle halo properties that are required include profiles describing the change in density as a function of radius within the halo, as well as a measure of the spin of the halo. This information is necessary to determine the effective rotational velocity of the halo, which will determine the angular momentum of any gas that cools and sinks towards the centre of the halo.

2.3.1 Halo density profile

The density profile of a halo is one of the fundamental descriptions of the internal structure of a halo and is necessary for calculating many other halo properties, including the enclosed mass, angular momentum, velocity dispersion and gravitational potential energy. The GALFORM model is set up to be able to adopt one of several different halo profiles.

By default, GALFORM assumes a (Navarro et al., 1995, hereafter NFW) density profile, $\rho(r)$, given by,

$$\rho(r) = \frac{\Delta_{\rm vir}\rho_{\rm crit}}{f(a_{\rm NFW})}\frac{1}{r/r_{\rm vir}(r/r_{\rm vir}+a_{\rm NFW})^2} \quad (r \leqslant r_{\rm vir}) \tag{2.11}$$

where the function $f(a_{\rm NFW}) = \ln(1+1/a_{\rm NFW}) - 1/(1+a_{\rm NFW})$ is truncated at the virial radius $r_{\rm vir}$, $a_{\rm NFW}$ is a scale-length that increases with halo mass (e.g. Cole & Lacey, 1996), $\rho_{\rm crit} = 3H^2/(8\pi G)$ is the critical density and $\Delta_{\rm vir}$ is the over-density, $\Delta = \bar{\rho}(< r)/\rho_{\rm crit}$. The virial radius is defined to be the radius at which the $\bar{\rho}(< r_{\rm vir}) = \Delta_{\rm vir}\rho_{\rm crit}$, where, for the spherical collapse model with $\Omega_0 = 1$, $\Delta_{\rm vir} = 178$. At small radii, the NFW profile tends towards a $\rho(r) \propto r^{-1}$ dependency, as was observed in simulations of dark matter halos at the time (Dubinski & Carlberg, 1991). At large radii, the profile tends towards the $\rho(r) \propto r^{-3}$ dependency. The profile was fitted to isolated halos in simulations carried out by Navarro et al.. Although the NFW profile has been shown to fit well the density profiles of many isolated halos in many N-body simulations with a wide range of halo masses and initial conditions (Navarro et al., 1996, 1997; Eke et al., 1998b; Frenk et al., 1999), many recent studies with improved simulations suggest that an Einasto (1965) profile is preferred (Navarro et al., 2004; Merritt et al., 2005; Prada et al., 2006).

2.3.2 Spin

During its formation, a dark matter halo will acquire angular momentum due to the action of tidal torques. Prior to turnaround, a growing perturbation will gain angular momentum in proportion to $(1+z)^{-3/2}$, as shown by White (1984). However, following collapse, the angular momentum of the halo will remain approximately constant (e.g. Zavala et al., 2008). The magnitude of the angular momentum of a halo is quantified by the dimensionless, *spin parameter*, $\lambda_{\rm H}$. The spin of a halo of mass $M_{\rm H}$ is defined as,

$$\lambda_{\rm H} = \frac{J_{\rm H}\,|E_{\rm H}|^{1/2}}{GM_{\rm H}^{5/2}}, \tag{2.12}$$

where $J_{\rm H}$ and $E_{\rm H}$ are the total angular momentum and energy of the halo within the virial radius. The total angular momentum of the mass of a halo within a radius, r, is given by,

$$J_{\rm H}(r) = \int_0^r \frac{\pi}{4} V_{\rm rot} r' \rho(r') 4\pi r'^2 {\rm d}r' \tag{2.13}$$

where $V_{\rm rot}$ is the halo rotation velocity. The total energy of the halo is equal to the sum of the potential and kinetic energies, $W_{\rm H}$ and $T_{\rm H}$ respectively. The total potential energy within a radius, r, is given by,

$$W_{\rm H}(r) = \frac{1}{2}\int_0^r \phi(r')\rho(r')4\pi r'^2 {\rm d}r', \tag{2.14}$$

where $\phi(r)$ is the gravitational potential, while the total kinetic energy is given by,

$$T_{\rm H}(r) = \frac{3}{2}\int_0^r \sigma^2(r')\rho(r')4\pi r'^2 {\rm d}r', \tag{2.15}$$

where $\sigma^2(r)$ is an isotropic velocity dispersion (assuming hydrostatic equilibrium). Note also how Eq. (2.13), Eq. (2.14) and Eq. (2.15) are dependent upon the density profile.

By default, GALFORM assigns the spin of a halo by drawing a value for $\lambda_{\rm H}$ from either a lognormal distribution (which has been shown to be a good approximation for the distribution of spins of halos extracted from cosmological N-body simulations, e.g. Cole & Lacey 1996) or from the distribution of Bett et al. (2007), which was obtained by fitting to the distribution of spins of the halos from the *Millennium Simulation*. As commented on by Benson & Bower (2010), the disadvantage of this statistical approach is that, firstly, the spin (and therefore angular momentum) is not influenced by the merger history of a halo and secondly, the spin can change dramatically between time-steps, even if the halo has not experienced any merging. As a result, evolution of the spin of a halo, and its dependence upon the merger history of the halo, is lost. Including such information is a necessary first

step if one wishes to model, for example, any intrinsic alignments between galaxies and their host halos, which are a possible systematic in gravitational lensing measurements. Benson & Bower suggest a scheme to follow the angular momentum evolution by, for each halo with multiple progenitors, using the spins of the progenitors to calculate their internal angular momenta and summing these momenta with statistically derived orbital momenta. They point out however that although this is a reasonable approach for mergers where one progenitor is significantly more massive, it fails for mergers between progenitors of equivalent mass. Additionally, material with high specific angular momentum could be ejected during merger events, leaving the descendant halo with a lower overall angular momentum. However, they provide a simple empirical correction that allows for good agreement with the spin distribution of the *Millennium Simulation* as measured by Bett et al. (2007). The main problem with measuring the spin of halos in simulations is that the halo must be resolved with many particles, typically $\gtrsim 200 - 300$, and so measurements are currently limited to the most massive halos.

2.4 Shock-heating and radiative cooling

For matter perturbations undergoing growth in the linear regime, the density field of baryonic matter (in the form of diffuse, cold gas) is thought to trace the density field of dark matter (Peacock, 1999). However, once sufficiently over-dense perturbations begin to undergo non-linear collapse and begin to form virialised (dark matter) halos, any gas located within the gravitational potential of the structure is thought to be shock heated to close to the virial temperature of the halo (Silk, 1977; Rees & Ostriker, 1977; White & Frenk, 1991). Over time, radiative processes lead to the gas losing energy and subsequently cooling, with the denser gas cooling more rapidly (due to the timescale for cooling being inversely proportional to the gas density). As this occurs the radiative pressure support, that would normally maintain hydrostatic equilibrium, is reduced and so cooling gas begins to sink in the gravitational potential of the halo. The notion of dissipative gas cooling within dark matter halos was first postulated by White & Rees (1978). The basic model of gas cooling, upon which GALFORM is based, was not presented in detail until the work of White & Frenk (1991).

2.4.1 Hot gas distribution

In the GALFORM model, any initially diffuse, cold gas located within the gravitational potential of a newly formed, virialised halo is assumed to be shock heated[3] to an isothermal temperature close to the virial temperature, $T_{\rm vir}$, for that halo, which is given by,

$$T_{\rm vir} = \frac{1}{2}\frac{\mu m_{\rm H}}{k_{\rm B}} V_{\rm H}^2, \tag{2.16}$$

where $m_{\rm H}$ is the mass of a hydrogen atom, μ is the mean molecular mass, $V_{\rm H}$ is the circular velocity of the halo and $k_{\rm B}$ is the Boltzmann constant.

By additionally assuming that this hot gas is still in hydrostatic equilibrium, the density profile of the hot gas follows,

$$\rho_{\rm gas}(r) = \rho_o \exp\left(-\frac{\Phi(r)}{c_T^2}\right), \tag{2.17}$$

where $\Phi(r)$ is the gravitational potential of the halo and $c_T^2 \equiv k_{\rm B}T/\left(\mu m_{\rm H}\right)$, where $T_{\rm gas} \sim T_{\rm vir}$ is the temperature of the gas. However, this simple solution yields an unphysical result at $r = 0$, where it predicts the density to diverge. By imposing boundary conditions, such as fixing the density at $r = 0$ to a constant value, ρ_0, and the gradient of the gravitational potential at $r = 0$ to zero, a physical solution can be obtained numerically. This solution is well approximated by a King (1972) profile,

$$\rho(r) = \frac{\rho_0}{\left[1 + (r/r_0)^2\right]^{3/2}} \tag{2.18}$$

where $r_0 = 3c_T/\sqrt{4\pi G\rho_0}$.

In GALFORM, the hot gas is assumed to settle into a cored density profile following the β-model of Cavaliere & Fusco-Femiano (1976),

$$\rho_{\rm gas} \propto \left(r^2 + r_{\rm core}^2\right)^{-3\beta_{\rm fit}/2} \tag{2.19}$$

where $r_{\rm core}$ is the radius of a central core (that is less concentrated than that of the dark matter) and $\beta_{\rm fit}$ is a free parameter. Fitting to typical cluster gas profiles from ΛCDM simulations by Navarro et al. (1995) and Eke et al. (1998a), suggests that $\beta_{\rm fit} \simeq 2/3$ and, initially, $r_{\rm core}/r_{\rm NFW} \simeq 1/3$, where $r_{\rm NFW} = a_{\rm NFW} r_{\rm vir}$.

As instigated by Cole et al. (2000), whenever a halo merger occurs, the hot gas in the descendant halo is assumed to again be shock heated back to close to the virial

[3]Gas accreted supersonically is assumed to be accretion shocked if the temperature of the gas is lower than the virial temperature of the halo onto which it is being accreted (Binney, 1977).

temperature of the new halo. Prior to the merger, however, the densest hot gas (located towards the centre of the halo, see Section 2.4) will have begun to cool. Since this gas will have contributed the lowest entropy, the remaining hot gas will have an increased minimum entropy. This increase in the minimum entropy over time will lead to an increase in the radius of the core in the gas density profile (Evrard & Henry, 1991; Kay & Bower, 1999; Wu et al., 2000). In GALFORM, if, following a merger, some fraction of the hot gas has cooled and the remaining fraction is smaller than the global gas fraction (equal to $\Omega_{\rm b}/\Omega_0$), then the core radius is increased until the gas density at the virial radius is equal to what we would expect if none of the gas had cooled. A limit of $r_{\rm core} \leqslant 10 r_{\rm vir}$ is set to avoid the situation where sufficient gas has cooled preventing the target density being attained. This limit, however, is rarely reached. Note, however, that in the Bower et al. (2006) GALFORM model, the radius of the core was kept fixed at its initial value.

The angular momentum of the cooling gas within a halo is determined by the rotational velocity of the halo. The mean rotational velocity of the hot gas, $V_{\rm rot}$, within concentric shells of material is assumed to remain constant with increasing radius and always has the same orientation. The rotational velocity is therefore found by,

$$V_{\rm rot} = A\left(a_{\rm NFW}\right) \lambda_{\rm H} V_{\rm H}, \tag{2.20}$$

where $V_{\rm H} = \left(GM/r_{\rm vir}\right)^2$ is the circular velocity of the halo at the virial radius and $A\left(a_{\rm NFW}\right)$ is a dimensionless coefficient displaying a weak dependence upon the $a_{\rm NFW}$ free parameter.

2.4.2 Gas cooling

Radiative, two-body processes, including Bremsstrahlung and, at lower temperatures, collisional ionisation, excitation and recombination, lead to a decrease in the internal energy of the hot gas component over time. The energy per unit mass of gas divided by the rate at which a unit mass of gas is radiating energy is used to define the *cooling time*, $\tau_{\rm cool}$. For a infinitely small shell of gas (assumed to be in collisional ionisation equilibrium) at a radius r from the halo centre, the cooling time is,

$$\tau_{\rm cool}(r) = \frac{3}{2} \frac{1}{m_{\rm H}\mu} \frac{kT_{\rm gas}}{\rho_{\rm gas}(r)\Lambda(T_{\rm gas}, Z_{\rm gas})} \tag{2.21}$$

where $\rho_{\rm gas}(r)$ is the gas density and $\Lambda(T_{\rm gas}, Z_{\rm gas})$ is the *cooling function*, which describes the cooling rate for a gas with temperature, $T_{\rm gas}$ and metallicity, $Z_{\rm gas}$. At high temperatures the cooling function tends towards $\Lambda\left(T_{\rm gas}, Z_{\rm gas}\right) \propto T^{1/2}$, as expected for ionised

plasmas emitting Bremsstrahlung radiation. At lower temperatures enriched gas with a higher metal content will emit radiation by various mechanisms, including Compton cooling, leading to a more complex temperature dependence. For metal poor, primordial gas, the cooling function typically has two peaks, one corresponding to the ionisation of hydrogen at a temperature of $15,000\,\mathrm{K}$ and the other corresponding to the ionisation of helium atoms at a temperature of $100,000\,\mathrm{K}$. Since the variation in the cooling rate with the temperature and metallicity of the gas is not straight-forward, GALFORM adopts, as input, tabulated cooling functions from Sutherland & Dopita (1993).

Since cooling in galactic and cluster mass halos is dominated by two-body processes, the cooling time is inversely proportional to the density of the gas and so the densest gas, located at smaller radii, cools fastest. As a result, $\tau_{\mathrm{cool}}(r)$ will increase with radius. It is therefore possible to define the *cooling radius*, r_{cool}, as the radius at which the cooling time of the gas is equal to the age of the halo, i.e. the radius at which $t = \tau_{\mathrm{cool}}$, where t is the time since the halo was formed. Thus, the cooling radius will propagate outwards with time and, as such, any hot gas located at $r \leqslant r_{\mathrm{cool}}(t)$ will be assumed to have had sufficient time to cool.

Since the cooling gas is no longer pressure supported, it will sink towards the centre of the halo where it is accreted onto a galactic disk. The time required for the gas to sink and be accreted, known as the *free-fall time*, t_{ff}, is given by

$$t_{\mathrm{ff}}(r) = \int_0^r \left[\int_r^{r''} -\frac{2GM\left(r'\right)}{r'^2} \mathrm{d}r' \right]^{-1/2} \mathrm{d}r''. \tag{2.22}$$

As before, it is possible to define a radius, known as the *free-fall radius*, r_{ff}, for which cooling gas will have had sufficient time to sink and be accreted onto the disk.

Just like the cooling radius, the free-fall radius will propagate outwards with time, though not necessarily at the same speed as the cooling radius. Depending upon which of these radii is larger, gas at a particular radius may have had sufficient time to cool but insufficient time to sink, or vice versa. As a result, in GALFORM, the mass of cold gas that is accreted onto the galactic disk, M_{cool}, is defined using the smallest of these two radii, i.e. $r_{\mathrm{min}}(t) = \min\left[r_{\mathrm{cool}}, r_{\mathrm{ff}}\right]$.

For each dark matter halo, identified at an epoch t_i, the cooling rate is calculated by determining the mass of gas that has cooled and been accreted since the immediately previous epoch, $t_{i-1} < t_i$, in the halo merger tree. This is done by calculating the effective cooling radius, $r_{\mathrm{min}}\left(t_i\right)$, for the halo and comparing this to the effective cooling radius for the most massive progenitor of the halo, $r_{\mathrm{min}}\left(t_{i-1}\right)$, located at the previous output in the

tree. The cooling rate is thus given by $\dot{M}_{\rm cool} = M_{\rm cool}/\left(t_i - t_{i-1}\right)$, where $M_{\rm cool}$ is the mass of gas located at radii $r_{\rm min}\left(t_{i-1}\right) < r \leqslant r_{\rm min}\left(t_i\right)$[4]. If the galaxy is a satellite in its host dark matter halo, then GALFORM assumes that the galaxy was stripped of its hot gas upon infall and so sets $\dot{M}_{\rm cool} = 0$. GALFORM uses the instantaneous recycling approximation (Tinsley, 1980) to model star formation, feedback and chemical enrichment as a set of differential equations (see Section 2.7), under the assumption that the flow of material between the three components is proportional to the instantaneous star formation rate, ψ, or the cooling rate, $\dot{M}_{\rm cool}$ (the latter of which is assumed to be constant over a timestep). The cooling rate is therefore an incredibly important quantity.

Since $\dot{M}_{\rm cool}$ is always computed using the initial density profile of the hot gas, any gas that has been reheated due to feedback (see Section 2.6) is assumed to not contribute until the halo undergoes a merger (i.e. a new formation event occurs) and this gas is incorporated back into the hot gas component. Bower et al. (2006) improved upon the cooling calculation by allowing reheated gas to cool back onto the galactic disk on a timescale comparable to the dynamical timescale of the halo, $\tau_{\rm dyn}$, which is often shorter than the timestep between two adjacent output times in the halo merger tree. If this timescale is $\tau_{\rm reheat} = \tau_{\rm dyn}/\alpha_{\rm reheat}$, where Bower et al. set $\alpha_{\rm reheat} \sim 1$, then the mass of reheated gas available for cooling is $\Delta M = M_{\rm reheat}\Delta t/\tau_{\rm reheat}$.

2.4.3 A note on shock validity and cold accretion

The "classical" cooling model assumed in semi-analytic models is that all gas gets shock heated to approximately the virial temperature of the halo and that cooling can occur in one of two distinct regimes (Binney, 1977; Rees & Ostriker, 1977; White & Frenk, 1991). GALFORM assumes that the shock occurs close to the virial radius of the halo. If the cooling time is shorter than the free-fall time of the gas, then the gas is accreted 'cold'. This is typical at high redshift, when the mean density of the Universe was higher, or in low mass halos at the present day. However, if the cooling time is longer than the free-fall time, as is typical in high-mass ($M_{\rm halo} \gtrsim 10^{11}h^{-1}{\rm M}_{\odot}$) halos assembled at late times, then a quasi-static halo of hot gas is formed. This picture has been supported by comparisons with the predictions from smooth particle hydrodynamics (SPH) simulations (Benson et al., 2001; Yoshida et al., 2002; Helly et al., 2003).

However, Birnboim & Dekel (2003) have recently argued that an accretion shock can

[4]Obviously, if the halo is located at the top of the tree and has no progenitor, then the $M_{\rm cool}$ is simply the mass of gas located at $r \leqslant r_{\rm min}\left(t_i\right)$.

only be supported by a stable atmosphere of hot gas and that shocks are only likely to occur close to the virial radius of a halo for high mass halos where a quasi-static hot halo has formed. For low mass halos, accretion shocks form at much smaller radii, closer to the scales of galaxy disks. It is worth noting that White & Frenk (1991) support this idea. However, whilst Birnboim & Dekel propose that whether a stable shock will be supported depends only on the properties of the accreted gas, White & Frenk argue that shock stability will depend on the accretion and cooling history of a halo, including possible reheating due to feedback mechanisms. Subsequent SPH simulations have suggested that a significant fraction of the gas in low mass galaxies has never undergone shock heating[5] (e.g. Fardal et al., 2001; Kereš et al., 2005; Ocvirk et al., 2008; Kereš et al., 2009).

Benson & Bower (2011) addressed this issue by introducing the Birnboim & Dekel stability criterion into a version of GALFORM. Despite the assumption made by GALFORM that gas is shock heated at the virial radius, Benson & Bower were able to mimic unstable shocks, and therefore cold accretion of the gas, by setting $\tau_{\rm cool} = 0$. The inclusion of this treatment of accretion shocks was found to have little impact on the high redshift cosmic star formation rate as well as on the galaxy luminosity function or the sizes of galactic disks in the local Universe. Instead, Benson & Bower find that uncertainties in the treatment of feedback from supernovae and active galactic nuclei have a much greater influence on the statistics of the galaxy population. From this they conclude that the "classical" model used in semi-analytic models is sufficiently accurate for the current applications of these models. Note, however, that the geometry of the gas accretion is different in the simulations, than is assumed in the semi-analytic models, with cooling occurring along filaments. This may have implications for the angular momentum of the accreted gas.

2.5 Star formation in disks

2.5.1 Dependence on cold gas mass

Prior to the work of (Lagos et al., 2011a, see Section 2.5.2), the assumption made in GALFORM was that the instantaneous star formation rate, ψ, in the disk of a galaxy was

[5]Note, however, that SPH simulations cannot adequately resolve shocks due to numerical artefacts (see, for example, Agertz et al., 2007) and so supporting evidence for a lack of shock heating is limited to regions that have sufficient resolution. Additionally, shocks do not happen in SPH simulations unless an artificial viscosity is introduced (e.g. Springel, 2010).

simply proportional to the mass of cold gas in the disk, $M_{\rm cold}$, i.e.

$$\psi = \frac{M_{\rm cold}}{\tau_\star} \tag{2.23}$$

where $\tau_\star$ is a star formation timescale.

Two different definitions for $\tau_\star$ have been adopted by different incarnations of GALFORM. Both definitions scale $\tau_\star$ with the circular velocity of the galactic disk according to a power-law. The first, adopted by all apart from Baugh et al. (2005), defines $\tau_\star$ as,

$$\tau_\star = \frac{\tau_{\rm disk}}{\varepsilon_\star} \left(V_{\rm disk}/200{\rm km\,s^{-1}}\right)^{\alpha_\star} \tag{2.24}$$

where $V_{\rm disk}$ is the circular velocity of the disk at the half-mass radius $r_{\rm disk}$, $\tau_{\rm disk} = r_{\rm disk}/V_{\rm disk}$ is the dynamical timescale of the disk and $\alpha_\star$ and $\varepsilon_\star$ are dimensionless, free parameters. The parameter $\varepsilon_\star$ determines the efficiency of star formation as the fraction of cold gas that is turned into stars per dynamical time for a galaxy like the Milky Way (i.e. with $V_{\rm disk} \simeq 200\,{\rm km\,s^{-1}}$), while $\alpha_\star$ is a power-law index. The values for $\varepsilon_\star$ and $\alpha_\star$ are constrained by the requirement that the GALFORM model be able to broadly reproduce a designated set of observational results (such as, for example, the trend of gas mass-to-light ratio with luminosity for spiral galaxies). This definition predicts star formation timescales that are broadly consistent with the low-redshift observations of Kennicutt (1998) (see also Bell et al., 2003).

The alternative definition, adopted by Baugh et al. (2005), retains the power-law scaling with disk circular velocity, but drops the dependence upon the disk dynamical timescale and instead adopts a simple, normalisation constant, $\tau_{\star 0}$:

$$\tau_\star = \tau_{\star 0} \left(V_{\rm disk}/200{\rm km\,s^{-1}}\right)^{\alpha_\star} . \tag{2.25}$$

Baugh et al. adopted this definition in order to attempt to reduce the amount of quiescent star formation activity at high redshift. Compared to the definition in Eq. (2.24), the definition in Eq. (2.25) leads to similar star formation timescales at $z \sim 0$, but much longer timescales at high redshift. The values of $\tau_{\star 0} = 8\,{\rm Gyr}$ and $\alpha_\star = -3$ were set such that the Baugh et al. model was able to match the gas mass-to-luminosity ratios of local galaxies (see Power et al., 2010) and to increase the number of starbursts at high redshift (to match the number counts of submillimetre galaxies and Lyman-break galaxies).

2.5.2 The Lagos et al. star formation recipe

The surface density of star formation rate in the disk of a galaxy, $\Sigma_{\rm SFR}$, is commonly set proportional to the surface density of gas in the disk, $\Sigma_{\rm gas}$ using either (i) the Schmidt

(1959) law, $\Sigma_{\rm SFR} \propto \Sigma_{\rm gas}^{N}$, where N is a power-law index, or (ii) the Shu (1973) law, $\Sigma_{\rm SFR} \propto \Sigma_{\rm gas}/\tau_{\rm dyn}$, where $\tau_{\rm dyn}$ is a dynamical timescale. Observational work by Kennicutt (1998) examining the correlations between the global star formation rate (SFR) and gas surface densities for a wide range of local galaxies, found that the global SFR relation is equally well fitted by using either a Schmidt law with $N = 1.40 \pm 0.15$ or a Shu law and setting $\tau_{\rm dyn}$ equal to the disk orbital time. However, observational studies have suggested a break in the power-law relation at low surface gas densities below a critical gas density (Kennicutt, 1989; Martin & Kennicutt, 2001).

Recent work by Wong & Blitz (2002); Kennicutt et al. (2007); Bigiel et al. (2008) supports $\Sigma_{\rm SFR}$ showing a linear dependence upon the surface density of molecular gas, $\Sigma_{\rm mol}$. This relation is much stronger than the relation between SFR surface density and the surface density of either the total cold gas or the atomic hydrogen content. As Lagos et al. (2011a) discuss, the observation that stars form in dense molecular clouds suggests that $\Sigma_{\rm SFR} \propto \Sigma_{\rm mol}$ is physically reasonable. Furthermore, star formation models by Blitz & Rosolowsky (2006) and Leroy et al. (2008), where the ratio of molecular-to-atomic hydrogen gas content is made proportional to a power-law function of the hydrostatic pressure in the disk, are able to reproduce a break in the $\Sigma_{\rm SFR} - \Sigma_{\rm mol}$ plane. However, the mechanism driving star formation in dense molecular clouds remains uncertain with various models having been suggested, including turbulence-driven molecule formation (Krumholz & McKee, 2005; Krumholz et al., 2009) and shearing of galactic disks causing cloud collisions (Tan, 2000; Schaye, 2004; Silk & Norman, 2009).

One possible star formation law, presented by Blitz & Rosolowsky (2006), is based upon two observationally motivated proposals. Firstly, that the surface density of SFR is linearly proportional to the molecular gas surface density,

$$\Sigma_{\rm SFR} = \nu_{\rm SF} \Sigma_{\rm mol}, \tag{2.26}$$

where $\nu_{\rm SF}$ is a proportionality factor given as an inverse timescale, and that, secondly, the ratio of molecular to atomic hydrogen gas, $R_{\rm mol}$, is set by a power-law relation with the internal hydrostatic pressure in the galactic disk, $P_{\rm hyd}$, i.e.,

$$R_{\rm mol} \equiv \frac{\Sigma\left(H_2\right)}{\Sigma\left(H_{\rm I}\right)} = \left(\frac{P_{\rm hyd}}{P_0}\right)^{\alpha}, \tag{2.27}$$

where $\log_{10}\left(P_0 k_{\rm B}^{-1}/{\rm K\,cm^{-3}}\right) = 4.54 \pm 0.07$ and $\alpha = 0.92 \pm 0.07$ are set by Blitz & Rosolowsky fitting Eq. (2.27) to the distribution of observed molecular-to-atomic ratios

from their galaxy sample. Hence, the Blitz & Rosolowsky star formation law is given by,

$$\Sigma_{\rm SFR} = \nu_{\rm SF}\Sigma_{\rm mol} = \nu_{\rm SF}\left(\frac{\Sigma_{\rm mol}}{\Sigma_{\rm gas}}\right)\Sigma_{\rm gas} = \nu_{\rm SF}\left(\frac{R_{\rm mol}}{R_{\rm mol}+1}\right)\Sigma_{\rm gas} \tag{2.28}$$

where $\nu_{\rm SF}$ is an inverse timescale and $\Sigma_{\rm gas}$ is the projected gas density in the galactic disk.

The inverse timescale, $\nu_{\rm SF}$, is described by,

$$\nu_{\rm SF} = \nu^0_{\rm SF}\left[1+\left(\frac{\Sigma_{\rm gas}}{\Sigma_0}\right)^q\right] \tag{2.29}$$

where $\nu^0_{\rm SF} = 0.525 \pm 0.25\,{\rm Gyr}^{-1}$, $\Sigma_0 = 200\,{\rm M}_\odot\,{\rm pc}^{-2}$ and $q = 0.4$, chosen to reproduce the Kennicutt (1998) star formation law at high gas densities and the steepening seen in the $\Sigma_{\rm SFR} - \Sigma_{\rm gas}$ relation by Bigiel et al. (2008).

In general, the instantaneous star formation rate is not simply proportional to the mass of cold gas mass in the disk and there is no simple analytic solution. Lagos et al. (2011a) therefore extended GALFORM to use fully numerically solved star formation recipes, so that more general star formation laws could be incorporated. Lagos et al. compared three different star formation prescriptions: the empirical Kennicutt-Schmidt law, the observationally motivated Blitz & Rosolowsky (2006) model and the theoretical model of Krumholz et al. (2009). By incorporating each prescription into the Baugh et al. (2005) and Bower et al. (2006) GALFORM models, Lagos et al. were able to examine which observational galaxy properties are capable of helping to distinguish between these different star formation laws. The different star formation laws were found to have little impact upon the universal star formation rate density and, therefore, have little affect upon the optical and near-infrared galaxy luminosity functions. They were, however, found to change the relative contribution between quiescent star formation and starbursts. However, the introduction of the Blitz & Rosolowsky and Krumholz et al. laws led to galaxies with lower cold gas masses, a lower cosmic mean cold gas density and therefore smaller gas-to-light ratios leading to a much better agreement with observations than was possible with the published Bower et al. (2006) GALFORM model. Furthermore, for the Bower et al. model, these laws lead to the development of a second, 'passive' sequence (in addition to the already known, 'active' sequence) in the SFR versus stellar mass plane, as seen in the distribution of SDSS galaxies in this plane (Brinchmann et al., 2004).

The Bower et al. GALFORM model, in combination with the Blitz & Rosolowsky law, was found to predict an HI mass function in best agreement with local observational estimates from Zwaan et al. (2005). As a result, this was adopted as the default Lagos

et al. version of GALFORM. It is worth noting that, unlike the prescriptions in Section 2.5.1, the Blitz & Rosolowsky star formation law has no free parameters (up to the uncertainties on the zero point and slope of the observationally constrained quantities).

2.6 Feedback mechanisms

Since the cooling time scales with the temperature of the gas, which scales (approximately) with the circular velocity of the host dark matter halo (and therefore the mass of the halo), we would expect cooling to be more efficient in less massive halos. Additionally, since at higher redshift the mean density of the Universe was higher (and, as discussed in Section 2.4.2, the cooling time is inversely proportional with density), then we would expect the effective cooling time to be shorter at higher redshift. Therefore, we would expect cooling to be much more efficient in small halos forming at high redshift (which will ultimately become the most massive halos, hosting galaxy clusters, that we see today). Without any physical mechanism to regulate the star formation in these halos, we might therefore expect the stellar mass density today, $\Omega_{\star,0}$, to be comparable to the baryon matter density, $\Omega_{\mathrm{b},0}$. However, observational estimates for the mass-to-light ratios of galaxies predict $\Omega_{\star,0} \ll \Omega_{\mathrm{b},0}$ (e.g. Cole et al., 2001). This problem, commonly known as the *overcooling problem* (e.g. White & Rees, 1978), can also be seen if we compare the luminosity function of galaxies to the halo mass function (scaled by a constant mass-to-light ratio). The galaxy luminosity function and scaled halo mass function have very different shapes and are a poor match (see, for example, Baugh, 2006). If the two functions are normalised around the break in the galaxy luminosity function, then the scaled halo mass function predicts far too many galaxies, both faintwards and brightwards of the break point, suggesting that the the efficiency with which stars form is scale dependent. For this to be the case, *feedback mechanisms*, preventing the formation of very massive, luminous galaxies (which are not seen in excess in the Universe today), are a necessary addition to any model of galaxy formation.

In GALFORM, heating of the cold gas in the galactic disk by winds and supernovae (SNe), photoionisation of the intergalactic medium (IGM) as well as heating and gas expulsion due to active galactic nuclei (AGN) are all considered as forms of feedback.

2.6.1 Supernova feedback and winds

An obvious way to prevent star formation is for the stellar population that is already in place to somehow reheat and eject cold gas from the galactic disk. Such feedback, thought to be caused by supernovae and stellar-driven (super-)winds, was originally invoked in galaxy formation models to match the shallow faint-end slope of the galaxy luminosity function (White & Rees, 1978; White & Frenk, 1991; Cole, 1991).

Assuming that the instantaneous star formation rate, ψ, remains constant with time over the duration of a supernova, then the total energy from all SNe explosions, $E(t)$, at a time, t, can be written,

$$E(t) = \mu_{\rm SN}\psi \int_0^t \epsilon\left(t'\right) \mathrm{d}t', \tag{2.30}$$

where $\mu_{\rm SN}$ is the number of SNe per unit mass of forming stars and $\epsilon\left(t'\right)$ is the total energy from a single supernova remnant with age t' (Dekel & Silk, 1986). In a simple argument, for a mass, $M_{\rm eject,SN}$, of cold gas to be ejected, the energy $E(t)$ must equal or exceed the binding energy of the gas, $M_{\rm eject}V_{\rm disk}^2/2$, where $V_{\rm disk}$ is the circular (escape) velocity of the disk. We can therefore see that the rate at which reheated gas is ejected, $\dot{M}_{\rm eject,SN}$ should be expressed as a function of the star formation rate and the disk velocity. In GALFORM, this rate is defined by,

$$\dot{M}_{\rm eject,SN} = \beta_{\rm SN}\left(V_{\rm disk}\right)\psi, \tag{2.31}$$

where $\beta_{\rm SN}\left(V_{\rm disk}\right)$ scales the efficiency of the SNe feedback with the circular velocity of the galactic disk. Note that for an energy-driven wind, $\beta_{\rm SN} \propto V_{\rm esc}^{-2}$, where $V_{\rm esc}$ is the escape velocity, whereas for a momentum-driven wind, momentum conservation implies that $\beta_{\rm SN} \propto V_{\rm esc}^{-1}$ is a more appropriate scaling. In GALFORM the scaling is parameterised following a power-law of the form,

$$\beta_{\rm SN}\left(V_{\rm disk}\right) = \left(V_{\rm disk}/V_{\rm hot}\right)^{-\alpha_{\rm hot}}, \tag{2.32}$$

where $V_{\rm hot}$ is a free parameter with units of velocity that normalises the strength of SNe feedback and $\alpha_{\rm hot}$ is a dimensionless power-law index. The parameters $V_{\rm hot}$ and $\alpha_{\rm hot}$ are constrained by comparison of the GALFORM predictions with a set of observational data, typically the present day optical and near-infrared galaxy luminosity functions. SNe feedback has the largest affect on low-mass galaxies, with the ejection of gas from the disks of such galaxies leading to a flattening of the faint-end slope of the luminosity function.

In an attempt to match the bright-end break of the near-infrared galaxy luminosity function, Benson et al. (2003) found it necessary to include a contribution to the SNe feedback in the form of high velocity winds, capable of expelling reheated gas out of the dark matter halo entirely. Such 'superwinds' have been inferred to exist from observations at both low and high redshift (Martin, 1999; Heckman et al., 2000; Pettini et al., 2002; Adelberger et al., 2003; Smail et al., 2003). The rate at which gas is ejected from the halo by superwinds, $\dot{M}_{\rm eject,SW}$, is defined in a similar way to mass ejection rate from SNe, i.e.

$$\dot{M}_{\rm eject,SW} = \beta_{\rm SW}\left(V_{\rm disk}\right)\psi, \tag{2.33}$$

where

$$\beta_{\rm SW}\left(V_{\rm disk}\right) = f_{\rm SW}\min\left[1, \left(V_{\rm disk}/V_{\rm SW}\right)^{-2}\right], \tag{2.34}$$

with parameter values set to $f_{\rm SW} = 2$ and $V_{\rm SW} = 200\,{\rm km\,s^{-1}}$ (Baugh et al., 2005). Unlike the cool gas that is reheated by SNe feedback, gas expelled by superwinds is not allowed to be accreted back onto any halos. Superwind feedback therefore attempts to suppress star formation in massive galaxies by decreasing the overall hot gas content of the halo, thereby increasing the cooling time. Although this leads to GALFORM being able to reproduce the bright-end break in the luminosity function, extremely energetic winds, with highly efficient energy coupling, are required. This led Benson et al. (2003) to suggest an additional contribution to the feedback energy budget, such as energy injection from matter accreting onto a supermassive black hole.

Note that by combining the contribution from both SNe and superwinds, the total mass ejection rate from the disk can be written,

$$\dot{M}_{\rm eject} = \beta\left(V_{\rm disk}\right)\psi = \left[\beta_{\rm SN}\left(V_{\rm disk}\right) + \beta_{\rm SW}\left(V_{\rm disk}\right)\right]\psi. \tag{2.35}$$

However, for subsequent GALFORM models that incorporate feedback from AGN, feedback from superwinds is removed by setting $f_{\rm SW} = 0$. In this case $\beta\left(V_{\rm disk}\right) = \beta_{\rm SN}\left(V_{\rm disk}\right)$ and $\dot{M}_{\rm eject} = \dot{M}_{\rm eject,SN}$.

2.6.2 Photoionisation

During the epoch of reionisation, high energy photons from stars and AGN hosted by the most massive halos will begin to reionise the largely primordial gas in the intergalactic medium (IGM). Towards the end of this epoch, once the IGM is almost fully ionised, the radiation from these sources will begin to penetrate into the gas inside less massive halos. The result is that the formation of galaxies in low-mass halos is suppressed. This is

due to cooling in these halos being prevented due to photoionisation and photoionisation heating. The absorption of high energy photons from a background radiation field leads to ionisation within the gas, thus changing the population densities of different ion species within the gas. The ionisation of the atomic hydrogen removes the hydrogen peak in the cooling function. This modification of the cooling function of the gas increases the cooling time. Further heating will occur when any energy from the photon, surplus to the necessary ionisation energy, is provided to the electron as kinetic energy. Additionally, the increased pressure from the ionised IGM can physically prevent baryons from falling into halos.

In GALFORM, feedback from photoionisation is implemented in the following way. Firstly, reionisation is assumed to happen instantly at a redshift, $z_{\rm reion}$. For all redshifts $z < z_{\rm reion}$, the cooling of gas is completely inhibited in halos with circular velocities less than some threshold value, $V_{\rm H} < V_{\rm crit}$. The value $V_{\rm crit} = 60\,{\rm km\,s^{-1}}$ is adopted, to conform with the results from the detailed modelling of the effect of the IGM on the cooling of gas in halos performed by Benson et al. (2002). However, Okamoto et al. (2008) have recently suggested that $V_{\rm crit} = 30\,{\rm km\,s^{-1}}$ is a more suitable value. The value assumed for $z_{\rm reion}$, is more flexible, with Baugh et al. (2005) assuming $z_{\rm reion} = 6$. However, later models have adopted a value of $z_{\rm reion} = 10$ that is more consistent with the recent WMAP measurements of the cosmic microwave background (Spergel et al., 2007; Dunkley et al., 2009).

2.6.3 Feedback from active galactic nuclei

Besides the necessity for a more feasible solution to the overcooling problem, two other observational phenomena had been plaguing early models of galaxy formation. The first was the apparent 'cosmic downsizing' (Cowie et al., 1996), whereby the most massive galaxies in the local Universe are observed to be old and red, with very little star formation, in complete contradiction to expectations from hierarchical galaxy formation. The second, is the absence of the cooling flows at the centre of rich galaxy clusters, where one would expect large quantities of gas to be undergoing cooling (Tamura et al., 2001; Peterson et al., 2003). Bower et al. (2006) argued that the same mechanism is almost certainly responsible for all of these phenomena and that the solution is the injection of energy from a central supermassive black hole. Since the observation of a tight correlation between the mass of the bulge of a galaxy, $M_{\rm bulge}$, and the mass of its central black hole, $M_{\rm BH}$ (Magorrian et al., 1998), there has been increased interest in how the properties of a

supermassive black hole and the properties of its host galaxy influence one another (Silk & Rees, 1998).

For an AGN to have an impact upon its host galaxy, the energy output from the AGN must be able to balance the binding energy of the galaxy. Mo et al. (2010) demonstrate this in a simple example. Assuming that the power available from mass accretion onto a black hole is $\dot{E}_{\rm acc} = \epsilon \dot{M}_{\rm BH} c^2$, where $\dot{M}_{\rm BH}$ is the change in the mass of the black hole, $M_{\rm BH}$, and ϵ is an efficiency factor, then by integrating this power over the age of the AGN one can express the total energy output as $E_{\rm acc} = \bar{\epsilon} M_{\rm BH} c^2$, where $\bar{\epsilon}$ is the mean efficiency. Next, using the virial theorem, an elliptical galaxy with mass, $M_{\rm gal}$, and velocity dispersion, σ, can be estimated to have a gravitational binding energy of $E_{\rm bind} \sim -M_{\rm gal}\sigma^2$. Thus,

$$\frac{E_{\rm acc}}{|E_{\rm bind}|} \sim \frac{\bar{\epsilon} M_{\rm BH}}{M_{\rm gal}} \left(\frac{c}{\sigma}\right)^2. \tag{2.36}$$

If we assume the result from Kormendy (2001), that $M_{\rm BH}/M_{\rm gal} \sim 10^{-3}$, then for a galaxy with $\sigma = 300\,{\rm km\,s^{-1}}$ we have $E_{\rm acc}/\left|E_{\rm bind}\right| \sim 1000\bar{\epsilon}$, suggesting that, even for modest efficiency factors, the energy output from an AGN can easily exceed the binding energy of a galaxy.

Feedback due to AGN was first implemented into the GALFORM model by Bower et al. (2006), who modelled AGN energy injection through a self-regulating feedback loop. Bower et al. assumed that AGN feedback will only occur in halos undergoing quasi-static cooling (as argued by Binney, 2004), that is, halos for which the cooling time of the hot gas at the cooling radius exceeds the free-fall time at the same radius, i.e.

$$\tau_{\rm cool}\left(r_{\rm cool}\right) > \frac{1}{\alpha_{\rm cool}} \tau_{\rm ff}\left(r_{\rm cool}\right), \tag{2.37}$$

where $\alpha_{\rm cool}$ is a free parameter with reasonable values close to unity. Bower et al. predict that Eq. (2.37) will be satisfied for halos with mass $M_{\rm halo} \gtrsim 2 \times 10^{11}\, h^{-1}{\rm M}_{\odot}$ and argue that the transition between a rapid cooling regime and a quasi-static hot halo provides a physical basis for the scale-dependent break in the galaxy luminosity function. Therefore, in GALFORM, for halos satisfying Eq. (2.37), the power of the AGN is calculated and compared with the energy radiated away by the flow of cooling gas. The growth of central black holes is modelled following Malbon et al. (2007) and Fanidakis et al. (2011). The AGN power is parameterised as a fraction, $\varepsilon_{\rm SMBH}$, of the Eddington luminosity of the black hole, $L_{\rm Edd}$. The Eddington luminosity is the power necessary for the outward radiation pressure on a shell of gas to balance the gravitational force acting upon the shell. Given by,

$$L_{\rm Edd} = \frac{4\pi G c}{\kappa} M_{\rm BH}, \tag{2.38}$$

where κ is the opacity of the gas[6], this is the only instance in GALFORM where the black hole mass enters into the calculation of the AGN power. Therefore, if

$$L_{\rm cool} < \varepsilon_{\rm SMBH} L_{\rm Edd}, \tag{2.39}$$

where $L_{\rm cool}$ is the luminosity of the cooling gas, then the hot gas is prevented from cooling and any cooling flows are quenched. The value of the $\varepsilon_{\rm SMBH}$ is set such that the model is able to reproduce the local $M_{\rm BH} - M_{\rm bulge}$ relation.

How supermassive black holes grow and the nature of the dominant mechanism coupling their energy output to the surrounding gas is still not certain. In GALFORM the dominant mode of black hole growth at high redshift is through the accretion of cold gas during galaxy mergers and disk instabilities (see Section 2.8.3). During this rapid accretion phase, AGN are thought to primarily transfer energy to the surrounding gas through radiation-driven processes, such as photoionisation (see Section 2.6.2) or momentum-driven winds. At late times, however, there is an increasing contribution from black holes accreting hot gas directly from the halo, as well as from black hole mergers (see Fanidakis et al. 2011). AGN in this 'hot-halo mode', typically have lower accretion rates and geometrically thick accretion disks. The build-up of magnetic fields in these thick disks is thought to generate powerful radio jets (see Fanidakis et al., 2011, for an interpretation of how the radio luminosity of AGN correlates with accretion rate, black hole mass and black hole spin), which have been observed to inflate giant X-ray cavities around elliptical galaxies at the centres of massive clusters (e.g. Forman et al., 2007). As such, radio jets are considered to be the key mechanism for the quenching of cooling flows by mechanical feedback.

The result of including AGN feedback in the Bower et al. (2006) model was that GALFORM was now able to successfully reproduce the break and bright-end slope of the luminosity function of local galaxies. Extending to higher redshift, the model was able to successfully predict the evolution of the galaxy stellar mass function out to $z \sim 5$ (Bielby et al., 2012) and, thanks to AGN feedback at late times, was able to explain the apparently antihierarchical behaviour of 'cosmic downsizing'. Additionally it was, for the first time, able to successfully reproduce the bimodal distribution of galaxy colours, as had been seen in galaxies from the SDSS (González et al., 2009), the clustering and abundance of luminous red galaxies as seen in the SDSS (Almeida et al., 2008) and match the number

[6]For ionised hydrogen, the opacity is equal to $\kappa = \sigma_{\rm T}/m_{\rm H}$, where $\sigma_{\rm T}$ is the Thomson scattering cross-section.

counts and redshift distribution of extremely red objects (Gonzalez-Perez et al., 2009).

2.7 Mass flow and chemical evolution

The cooling of gas, star formation (and subsequent stellar evolution) and feedback mechanisms are all fundamental processes dictating the evolution of the galaxy and the transfer of mass (including metals) between the three main components of the galaxy-halo system: the hot and cold gas components and the stellar population. In GALFORM the exchange of mass and metals between these different components is described using a series of differential equations. However, before considering these equations, it is useful to introduce the concept of the initial mass function.

2.7.1 The initial mass function

Since the evolutionary paths of stars are primarily determined by their mass, a crucial ingredient in the modelling of the evolution of a stellar population is the stellar *initial mass function* (IMF). In any star formation event, the IMF, $\phi(M)$, governs the initial number of stars formed as a function of mass, M, per unit stellar mass. It is typically modelled by a power-law,

$$\phi(M) \equiv \frac{\mathrm{d}N}{\mathrm{d}\ln M} \propto M^{-x}, \tag{2.40}$$

where x is the power-law index. Various choices of IMF exist, each with different mass-dependent values of x (e.g. Salpeter, 1955; Kroupa, 2001; Chabrier, 2003). The best constraints on the IMF usually come from observations of the solar neighbourhood. The default IMF adopted in GALFORM is the Kennicutt (1983) IMF where the power-law index is described by,

$$x = \begin{cases} 0.4, & \text{for } 0.15 < M/\mathrm{M}_\odot < 1 \\ 1.5, & \text{for } 1 < M/\mathrm{M}_\odot < 125 \end{cases}. \tag{2.41}$$

Mass limits such as those above are typically chosen since objects with masses below $\sim 0.1\,\mathrm{M}_\odot$ have central temperatures that are too cool to initiate hydrogen fusion, while objects with masses above $\sim 100\,\mathrm{M}_\odot$ are too unstable. For almost all GALFORM models the Kennicutt IMF is considered universal and used to describe both quiescent star formation and starbursts. The exception, however, is the Baugh et al. (2005) model which uses the Kennicutt IMF to describe quiescent star formation, but has starbursts governed by a top-heavy IMF, where $x = 0$ for all masses. Baugh et al. found it necessary to introduce the top-heavy IMF in bursts in order to match the number counts and redshift distributions

of faint submillimetre galaxies. There is much observational evidence in support of a top-heavy IMF in environments of intense star formation (e.g. Rieke et al., 1993; Larson, 1998; Figer et al., 1999; McCrady et al., 2003; Stolte et al., 2005; Larson, 2005; Fardal et al., 2007; Paumard et al., 2006; Parra et al., 2007; van Dokkum, 2008).

2.7.2 Flow of mass

In GALFORM, the flow of mass and metals is described under the assumption of the *instantaneous recycling approximation* (Tinsley, 1980), where, following an episode of star formation, mass and metals are returned to the interstellar medium instantaneously. For stellar populations with ages of a few Gyrs, this is a reasonably fair assumption (Nagashima et al., 2005). A consequence of this is that GALFORM is able to describe the mass flux between the stars and the hot and cold gas phases using a parameterisation that depends principally on the instantaneous star formation rate, ψ, and the cooling rate, $\dot{M}_{\rm cool}$. (Details of how these properties are calculated can be found in Sections 2.5 and 2.4.2 respectively.)

The flow of mass into the stellar component (i.e. the change in the mass locked up in stars), $\dot{M}_\star$, is equal to,

$$\dot{M}_\star = (1-R)\psi, \tag{2.42}$$

where R is the fraction of mass that is recycled by stars, i.e. the stellar mass that is lost due to winds and SNe. In reality, the fraction of material will change with time, t, and metallicity, Z, according to,

$$R(t;Z) = \int_{M(t;Z)}^{\infty} \left[M - M_r(M;Z)\phi(M)\right] \frac{{\rm d}M}{M} \tag{2.43}$$

where $M_r(M)$ is the remnant mass of a star with initial mass M. However, under the instantaneous recycling approximation, the recycled fraction is assumed to remain constant, i.e. $R(t;Z) \rightarrow R$. The appearance of the IMF in Eq. (2.43) means that the recycled fraction is not a free-parameter of the model, but can be fixed following selection of an IMF.

The flow of mass into the hot gas phase, $\dot{M}_{\rm hot}$, is given by,

$$\dot{M}_{\rm hot} = -\dot{M}_{\rm cool} + \beta\psi, \tag{2.44}$$

where β is the efficiency of feedback, introduced in Section 2.6.1. This flux is simply the sum of the mass of hot gas lost to cooling and the mass of cold gas regained due to SNe.

Recall that we have already seen in Eq. (2.31) that the second term, $\beta\psi$, is equal to the mass of cold gas ejected by SNe explosions.

Finally, the change in the mass of cold gas, $\dot{M}_{\rm cold}$, is expressed as,

$$\dot{M}_{\rm cold} = \dot{M}_{\rm cool} - (1 - R + \beta)\psi, \tag{2.45}$$

which is simply the mass of gas gained from cooling, minus the sum of the mass of cold gas locked up in stars and the mass of cold gas ejected from the disk due to SNe, i.e. $\dot{M}_{\rm cool} - \left(\dot{M}_{\star} + \dot{M}_{\rm eject}\right)$. Note that the action of AGN feedback in the Bower et al. (2006) GALFORM model is to simply set $\dot{M}_{\rm cool} = 0$.

The action of winds and SNe feedback, combined with ongoing star formation, leads to the cold gas reservoir in the disk of the galaxy being depleted on a timescale that is typically faster than $\tau_{\star}$. As a result, Cole et al. (2000) defined the true, effective timescale as $\tau_{\rm eff} = \tau_{\star}/(1 - R + \beta)$.

2.7.3 Chemical enrichment

The chemical enrichment of a galaxy-halo system is a crucial part of the galaxy formation process. The first stars born in the Universe, known as Population III stars, formed from primordial gas, which was predominantly hydrogen, with a small fraction of helium and a negligible metal content. However, the process of nuclear fusion within a star leads to a build-up of metals (heavy elements beyond hydrogen and helium) that will be dispersed in the interstellar medium and intergalactic medium through SNe explosions and stellar winds. As a result, the hot gas in the halo and the cold gas in the galactic disk, as well as later generations of stellar populations, will become increasingly polluted with metals and thus change the properties of the gaseous and stellar components. This change in the chemical composition will, for example, modify the cooling function and therefore affect the cooling time of the hot gas. Additionally, the fusing of heavy elements into dust grains will absorb short wavelength radiation (typically UV and optical wavelengths) and re-radiate it at longer wavelengths (typically infrared and submillimetre wavelengths). This extinction due to dust will therefore affect our measurements of the observational properties of galaxies.

Just like the mass flux, GALFORM is able to incorporate the production of metals and model their distribution throughout the three component reservoirs. If $Z_{\rm cold} = M^{Z}_{\rm cold}/M_{\rm cold}$ and $Z_{\rm hot} = M^{Z}_{\rm hot}/M_{\rm hot}$ are the mass fractions of cold gas and hot gas that are in the form of metals, then the differential equations governing the flow of metals can

be written as,

$$\dot{M}_{\star}^{Z} = (1-R) Z_{\rm cold} \psi, \tag{2.46}$$

$$\dot{M}_{\rm hot}^{Z} = -\dot{M}_{\rm cool} Z_{\rm hot} + (pe + \beta Z_{\rm cold})\, \psi, \tag{2.47}$$

$$\dot{M}_{\rm cold}^{Z} = \dot{M}_{\rm cool} Z_{\rm hot} + [p(1-e) - (1+\beta-R) Z_{\rm cold}]\, \psi, \tag{2.48}$$

where p is the *yield*, i.e. the fraction of mass converted into stars that is returned to the interstellar medium in the form of metals and e is the fraction of newly produced metals that are ejected from the galactic disk directly to the hot gas reservoir (typically $e = 0$). Just like the recycled fraction, the yield, $p_i(t; Z)$, for an element i will actually change as a function of time according to,

$$p_i(t; Z) = \int_{M(t:Z)}^{\infty} M_i\left(M_0; Z\right) \phi\left(M_0\right) \frac{{\rm d}M_0}{M_0}, \tag{2.49}$$

where $M_i\left(M_0; Z\right)$ is the mass of element i produced by stars with an initial mass M_0. Again, under the assumption of the instantaneous recycling approximation, once an IMF has been adopted and the yields calculated for all chemical elements of interest, the value for the yield can be assumed to be constant, $p(t; Z) \rightarrow p$.

2.8 Mergers and spheroid formation

Within the ΛCDM framework, dark matter halos are assumed to grow hierarchically through mergers. When this occurs, the galaxies hosted by the two progenitor halos will form a system of satellite galaxies orbiting around a central galaxy. The assumption made in GALFORM is that the central galaxy of the descendant halo is taken to be the central galaxy from the most massive progenitor. This is the case when working with either Monte-Carlo or N-body halo merger trees. Note that according to this definition, the central galaxy need not be the most massive or the most luminous galaxy in the halo. Over time, dynamical friction causes the satellite galaxies to lose energy and angular momentum, leading to a decay of their orbits and an eventual merger between the central galaxy and each satellite. Within GALFORM, galaxy mergers and the development of dynamically unstable galaxy disks are considered to be the only mechanisms capable of forming spheroids.

2.8.1 Positioning of galaxies in N-body halos

When working with N-body halo merger trees, after the merger of two halos the positions of galaxies within the descendant halo need to be considered. In this situation the central

galaxy is simply placed at the centre of mass of the most massive sub-group.

Satellite galaxies are placed at the centre of mass of the main halo sub-group originating from the progenitor halo in which that galaxy was last a central galaxy. If this sub-group no longer exists and it does not have a descendant sub-group, then the galaxy is simply placed on what was the most bound particle from that sub-group.

2.8.2 Dynamical friction

Following the merger of two halos, any infalling satellite galaxies will experience an apparent drag force, known as dynamical friction, which will cause them to lose kinetic energy and momentum. This effect can be visualised by considering a single, sufficiently massive particle passing through a large background of less massive, collisionless 'field' particles. Over time, numerous interactions between particle pairs will leave each particle in the system with the same mean kinetic energy (as the system re-establishes thermodynamic equilibrium). Since the infalling particle will tend to have a higher initial kinetic energy, it will therefore tend to lose energy to the field particles and so be slowed down. An alternative picture is to consider dynamical friction as being due to the gravitational force on the infalling particle due to an over-density of field particles that builds up behind it. Dynamical friction will cause the orbit of a satellite galaxy to decay, leading to the satellite galaxy merging with the central galaxy.

According to Lacey & Cole (1993), the estimated time taken for an orbit inside an isothermal halo to decay, $\tau_{\rm mrg}$, is given by,

$$\tau_{\rm mrg} = f_{\rm df}\Theta_{\rm orbit}\tau_{\rm dyn}\frac{0.3722}{\ln\left(M_{\rm H}/M_{\rm sat}\right)}\frac{M_{\rm H}}{M_{\rm sat}}, \tag{2.50}$$

where $M_{\rm H}$ is the mass of the halo inside which the satellite is orbiting, $M_{\rm sat}$ is the combined mass of the satellite galaxy and the halo inside which it formed, $\tau_{\rm dyn} \equiv \pi r_{\rm vir}/V_{\rm H}$ is the dynamical time of the descendant halo, $\Theta_{\rm orbit}$ is a factor incorporating the initial energy and angular momentum of the satellite orbit and $f_{\rm df}$ is a dimensionless parameter describing the efficiency with which the progenitor halo of the infalling satellite is stripped. The orbital factor, $\Theta_{\rm orbit}$, is defined by,

$$\Theta_{\rm orbit} = \left[J/J_{\rm c}(E)\right]^{0.78}\left[r_{\rm c}(E)/r_{\rm vir}\right]^2, \tag{2.51}$$

where E and J are the initial energy and angular momentum of the orbit of the satellite and $r_{\rm c}(E)$ and $J_{\rm c}(E)$ are the radius and angular momentum of a circular orbit with an energy equal to the initial energy of the satellite. The form of Eq. (2.51) was determined

by fitting to analytic results from Lacey & Cole (1993). Following on from Tormen (1997), Cole et al. (2000) conclude $\Theta_{\rm orbit}$ to be adequately described by a log-normal distribution.

In GALFORM, when a new halo forms, a satellite galaxy is assumed to enter the halo on a random orbit and have a random value for $\Theta_{\rm orbit}$ drawn from the log-normal distribution described previously. Alternatively, in the case where GALFORM is working with N-body halo merger trees, it is possible to use the position and velocity of the sub-group in which the galaxy is located prior to the merger, to calculate the value of $\Theta_{\rm orbit}$. Thus, for each satellite a value for $\tau_{\rm mrg}$ is calculated. After this time interval has elapsed the satellite is assumed to have merged with the central galaxy. If, however, the halo undergoes another merger before this time, then a new value for $\tau_{\rm mrg}$ is calculated based upon the orbit of the satellite in the new, more massive halo.

2.8.3 Galaxy mergers

In GALFORM, when two galaxies merge, the outcome of the merger depends upon the ratio of the mass (stars plus cold gas) of the satellite galaxy, $M_{\rm sat}$, to that of the central galaxy, $M_{\rm cen}$. (Note that, in the event of more than two galaxies merging simultaneously, the satellite galaxies are ranked by mass and merged sequentially with the central in a series of galaxy-galaxy mergers.)

If the mass ratio of the merging galaxies is $M_{\rm sat}/M_{\rm cen} \geqslant f_{\rm ellip}$, where $f_{\rm ellip}$ is a threshold parameter, then the merger is deemed to be *major merger*, the result of which is the formation of a single spheroid or elliptical galaxy. In this instance, all of the cold gas in the merging galaxies is converted into stars in a starburst event, on a timescale $\tau_{\star\rm burst} = \max\left[f_{\rm dyn}\tau_{\rm dyn}, \tau_{\rm burst,min}\right]$, where $\tau_{\rm dyn}$ is the dynamical time of the new spheroid and $f_{\rm dyn}$ and $\tau_{\rm burst,min}$ are free parameters to be constrained by fitting the model predictions to observational data.

Alternatively, if the mass ratio is $M_{\rm sat}/M_{\rm cen} < f_{\rm ellip}$, then the merger is deemed to be a *minor merger*. In this instance the galactic disk of the central galaxy is preserved and the stars of the satellite galaxy are added to a bulge component of the central galaxy, with any accreted cold gas from the satellite galaxy being added to the cold gas in the disk of the central galaxy (without changing the specific angular momentum of the disk). Baugh et al. (2005) updated GALFORM by allowing some minor mergers to undergo starbursts, provided that the galaxy mass ratio exceeds a threshold value $f_{\rm burst}$ (i.e. $f_{\rm burst} < M_{\rm sat}/M_{\rm cen} < f_{\rm ellip}$) and that the gas fraction of the central galaxy exceeds a critical value of $f_{\rm gas,crit} = M_{\rm cold}/M_{\star}$. The latter requirement is motivated by the suggestion that gas-rich disks are

more susceptible to starbursts triggered by the accretion of minor satellites (Hernquist & Mihos, 1995).

For all published models of GALFORM, the value of $f_{\rm ellip} = 0.3$ is adopted. Regarding starbursts occurring in minor mergers, Baugh et al. (2005) assume $f_{\rm burst} = 0.05$ and $f_{\rm gas,crit} = 0.75$, while for the Bower et al. (2006) model (and all subsequent models) $f_{\rm burst} = 0.1$ and $f_{\rm gas,crit} = 0.1$ are adopted.

2.8.4 Disk instabilities

Sufficiently massive galactic disks, with strong self-gravity, will become unstable to small perturbations from orbiting minor satellites or dark matter substructures. Such perturbations can lead to the formation of a galactic bar, which is capable of redistributing the mass and angular momentum of the disk. After a few dynamical times, sufficient mass can migrate towards the centre of the disk leading to a dense concentration and the formation of a central spheroid.

Following work by Efstathiou et al. (1982), galactic disks in GALFORM are considered unstable if,

$$\varepsilon_{\rm m} \equiv \frac{V_{\rm disk}}{\left(GM_{\rm disk}/r_{\rm disk}\right)^{1/2}} < 1.1, \tag{2.52}$$

where $V_{\rm disk}$ is the disk circular velocity at the disk half-mass radius, $M_{\rm disk}$ is the mass of the disk and $r_{\rm disk}$ is the half-mass radius of the disk. Any unstable disks are assumed to form a bar and, subsequently, a spheroid. In this case the stellar mass of the disk is transferred to the spheroid and any gas undergoes a starburst. Following the incorporation of AGN feedback by Bower et al. (2006), a fraction, $F_{\rm BH}$, of the disk gas is assumed to be accreted onto the central black hole. Bower et al. adopt $F_{\rm BH} = 0.005$ for all starbursts, irrespective of whether they were caused by galaxy mergers or by disk instabilities.

2.9 Galaxy sizes

As we have seen in previous sections, many galaxy properties display some dependence upon the size of the galactic disk, meaning that getting the sizes of disks (and spheroids) correct is essential. With a knowledge of the specific angular momentum and total mass of the gas that cools to form a galactic disk, the size of the resulting galactic disk can be determined. For calculating the size of spheroids, formed through either galaxy mergers or disk instabilities, GALFORM assumes that the galaxies are in virial equilibrium and that the binding energy of the descendant spheroid is equal to the total energy of the merging

galaxies, or, in the case of disk instabilities, the unstable disk and any pre-existing bulge component.

2.9.1 Halo contraction

The condensation of cold gas onto a galactic disk will enhance the gravitational potential at the centre of the dark matter halo, causing the halo to be compressed and dark matter particles to move towards the centre.

In GALFORM, the specific angular momentum, $rV_{\rm c}(r)$, of a particle located at a radius, r, with circular velocity $V_{\rm c}(r)$, is taken to be adiabatically invariant. This is a reasonable assumption if the condensation of the central galaxy occurs sufficiently slowly. Therefore, for a shell of dark matter at an initial radius, r_0, and with an initial circular velocity, $V_{\rm c,H}^{(0)}(r_0)$, we can write,

$$r_0 V_{\rm c,H}^{(0)}(r_0) = rV_{\rm c,H}(r), \tag{2.53}$$

where $V_{\rm c,H}(r)$ is the circular velocity of the shell at the final radius, r, of the shell. If the baryons and dark matter in the halo share a similar mass distribution and a fraction $1 - f_{\rm H}$ of the total mass goes into forming the galaxy, then the final, $M_{\rm H}(r)$, and initial, $M_{\rm H0}(r_0)$, masses contained within the shell are related by

$$M_{\rm H}(r) = f_{\rm H} M_{\rm H0}(r_0)\,. \tag{2.54}$$

By further assuming that the matter distributions are spherically symmetric, the final circular velocity of the system is,

$$V_{\rm c,H}^2(r) = G\left[M_{\rm H}(r) + M_{\rm D}(r) + M_{\rm B}(r)\right]/r \tag{2.55}$$

where $M_{\rm D}(r)$ and $M_{\rm B}(r)$ are, respectively, the masses of the galactic disk and bulge within the radius r. Incorporating the initial halo mass from Eq. (2.54) and making use of the equality in Eq. (2.53), gives,

$$r_0 M_{\rm H0}(r_0) = r\left[f_{\rm H} M_{\rm H0}(r_0) + M_{\rm D}(r) + M_{\rm B}(r)\right], \tag{2.56}$$

which relates, for given disk and bulge mass profiles, the final radius of any halo shell, following contraction, to its initial radius. From this the rotation curve of the halo can be computed, with the influence of the galaxy incorporated into the calculation.

2.9.2 Sizes of disk galaxies

The size of the galactic disk is basically governed by the angular momentum of the cool gas from which it forms. In GALFORM, the size of the disk is solved for by assuming conservation of angular momentum and centrifugal equilibrium.

Galactic disks in GALFORM are assumed to follow exponential surface density profiles,

$$\Sigma_{\rm D} = \frac{M_{\rm D}}{2\pi h_{\rm D}^2} \exp\left(-\frac{r}{h_{\rm D}}\right), \tag{2.57}$$

where $M_{\rm D}$ is the total mass of the disk and $h_{\rm D}$ is the radial scale length of the disk, which is related to the disk half-mass radius, $r_{\rm D}$, by $2\left(r_{\rm D}/h_{\rm D}+1\right)\exp\left(-r_{\rm D}h_{\rm D}\right)$. Integrating $\Sigma_{\rm D}(r)$ allows one to find the mass of the disk enclosed within a radius r,

$$M_{\rm D}(r) = \int_0^r 2\pi r'\Sigma_{\rm D}\left(r'\right){\rm d}r' = M_{\rm D}\left[1-\left(1+r/h_{\rm D}\right)\exp\left(-r/h_{\rm D}\right)\right]. \tag{2.58}$$

The specific angular momentum of the disk, $j_{\rm D}$, is given by the integral,

$$j_{\rm D} = \int_0^\infty 2\pi r\Sigma_{\rm D}(r) r V_{\rm cD}(r){\rm d}r/M_{\rm D} = k_{\rm D} r_{\rm D} V_{\rm cD}(r) \tag{2.59}$$

where $V_{\rm cD}(r)$ is the circular velocity profile of the disk and

$$k_{\rm h} = \int_0^\infty \frac{r^2}{h_{\rm D}^2}\exp\left(-\frac{r}{h_{\rm D}}\right)\frac{V_{\rm cD}(r)}{1.68V_{\rm cD}\left(r_{\rm D}\right)}\frac{{\rm d}r}{h_{\rm D}}. \tag{2.60}$$

Adopting a flat rotation curve leads to $k_{\rm D} = 1.19$. However, Cole et al. (2000) argue that the value of $k_{\rm D}$ is only weakly dependent on the form of the rotation curve, so this value is adopted for all GALFORM galaxies. Using $V_{\rm cD}^2 = GM\left(r_{\rm D}\right)/r_{\rm D}$, where $M\left(r_{\rm D}\right)$ is the total mass of dark matter and baryons within $r_{\rm D}$, and noting that by definition $M_{\rm D}\left(r_{\rm D}\right) = M_{\rm D}/2$, then we have,

$$j_{\rm D}^2 = k_{\rm D}^2 G r_{\rm D}\left[f_{\rm H}M_{\rm H0}\left(r_{\rm D0}\right) + \frac{1}{2}k_{\rm h}M_{\rm D} + M_{\rm B}\left(r_{\rm D}\right)\right], \tag{2.61}$$

where $k_{\rm h}$ is a constant that accounts for the fact that the disk is not spherically symmetric. Binney & Tremaine (1987) find that for an exponential disk, $k_{\rm h} = 1.25$. Since the disk half-mass radius, $r_{\rm D}$, must satisfy Eq. (2.61) and Eq. (2.56) evaluated at $r_{\rm D}$,

$$r_0 M_{\rm H0}\left(r_{\rm D0}\right) = r_{\rm D}\left[f_{\rm H}M_{\rm H0}\left(r_{\rm D0}\right) + M_{\rm D}\left(r_{\rm D}\right) + M_{\rm B}\left(r_{\rm D}\right)\right] = \frac{j_{\rm D}^2}{k_{\rm D}^2 G} - \frac{1}{2}\left(k_{\rm h}-1\right)r_{\rm D}M_{\rm D}, \tag{2.62}$$

then we can numerically solve this pair of coupled equations to obtain $r_{\rm D}$.

The size of the bulge (spheroid) component of the galaxy can be determined by setting up a similar set of coupled equations. The bulge is assumed to have a projected density

profile following a de Vaucoulers' $r^{1/4}$ law, which is characterised by an effective radius, $r_{\rm eff}$, defined as the radius containing half of the mass in projection. This is related to the three-dimensional half-mass radius of the bulge, $r_{\rm B}$, by $r_{\rm B} = 1.35 r_{\rm eff}$. A pseudo-specific angular momentum, assumed to be conserved, can be defined as $j_{\rm B} = r_{\rm B} V_{\rm c}\left(r_{\rm B}\right)$, where $V_{\rm c}(r)$ is the circular velocity profile within the bulge. As with the disk radius, we can relate the bulge radius and the specific angular momentum using the equations,

$$j_{\rm B}^2 = G r_{\rm B} \left[f_{\rm H} M_{\rm H0}\left(r_{\rm B0}\right) + M_{\rm D}\left(r_{\rm B}\right) + \frac{1}{2} M_{\rm B} \right] \tag{2.63}$$

and

$$r_{\rm B0} M_{\rm H0}\left(r_{\rm B0}\right) = \frac{j_{\rm B}^2}{G}, \tag{2.64}$$

which can be solved numerically to determine the bulge radius.

2.9.3 Sizes of spheroids from mergers

The assumption made by GALFORM is that two galaxies have merged when their separation equals the sum of their half-mass radii. If all of the galaxies (both the merging progenitors and the descendant) are in virial equilibrium, then their total energy is equal to half of their gravitational binding energy, i.e.

$$E_{\rm bind} = -\frac{c}{2} \frac{GM}{r} \tag{2.65}$$

where M and r are the mass and half-mass radius of the galaxy respectively and c is a form-factor related to the distribution of mass within the galaxy. Binney & Tremaine (1987) determine $c = 0.45$ for a $r^{1/4}$ profile and $c = 0.49$ for an exponential disk, though, for simplicity, GALFORM adopts $c = 0.5$ for all galaxies. If, just before they merge, the orbital energy of two galaxies, with masses M_1 and M_2 and half-mass radii r_1 and r_2, is

$$E_{\rm orbit} = -\frac{f_{\rm orbit}}{2} \frac{G M_1 M_2}{r_1 + r_2}, \tag{2.66}$$

where the parameter $f_{\rm orbit}$ depends upon the orbital parameters of the pair of merging galaxies, then, by energy conservation, the binding energy of the descendant spheroid will be $E_{\rm bind,sph} = E_{\rm bind,1} + E_{\rm bind,2} + E_{\rm orbit}$. GALFORM adopts $f_{\rm orbit} = 1$ by modelling the galaxies as point masses on circular orbits about a common centre of mass. Therefore, the masses and radii of the galaxies can be related through the equation,

$$\frac{\left(M_1 + M_2\right)^2}{r_{\rm sph}} = \frac{M_1^2}{r_1} + \frac{M_2^2}{r_2^2} + \frac{f_{\rm orbit}}{c} \frac{M_1 M_2}{r_1 + r_2}, \tag{2.67}$$

where $r_{\rm sph}$ is the half-mass radius of the descendant spheroid. This value of $r_{\rm sph}$, along with the mass of the spheroid, $M_{\rm sph} = M_1 + M_2$ can then be used to determine a value for $j_{\rm B}$ following the calculations in Section 2.9.2, so that the size of the spheroid can be determined with the effect of halo contraction having been properly taken into account.

2.9.4 Sizes of spheroids from disk instabilities

The sizes of spheroids formed as a result of disk instabilities are calculated in a similar way to those formed from galaxy mergers. For an unstable disk with mass $M_{\rm disk}$ and half-mass radius $r_{\rm disk}$, the radius, $r_{\rm sph}$, of the resulting spheroid is determined from

$$\frac{c_{\rm B}\left(M_{\rm disk}+M_{\rm bulge}\right)^2}{r_{\rm sph}} = \frac{c_{\rm D}M_{\rm disk}^2}{r_{\rm disk}} + \frac{c_{\rm B}M_{\rm bulge}^2}{r_{\rm bulge}} + f_{\rm int}\frac{M_{\rm disk}M_{\rm bulge}}{r_{\rm disk}+r_{\rm bulge}}, \qquad (2.68)$$

where $M_{\rm bulge}$ and $r_{\rm bulge}$ are the mass and half-mass radius of any pre-existing bulge and $c_{\rm D}$ and $c_{\rm B}$ are form factors relating to the distribution of mass within the disk and bulge respectively. For an exponential disk $c_{\rm D} = 0.49$, whilst any bulge component has $c_{\rm B} = 0.45$ (corresponding to the value for a $r^{1/4}$ profile). The last term in Eq. (2.68) describes the gravitational interaction energy between the disk and any bulge component. In GALFORM a value of $f_{\rm int} = 2$ is adopted. As before, the size and mass of the spheroid are then used to calculate the pseudo-specific angular momentum of the spheroid and update the spheroid radius following the effects of halo contraction.

2.10 Observable galaxy properties

The majority of the features of the GALFORM model discussed in the previous sections have so far been concerned with many of the intrinsic properties of galaxies (including the star formation history, chemical evolution and size evolution) that cannot be easily measured observationally. Therefore, in order to compare the model predictions with observations it is necessary to predict the observable properties of galaxies, such as their luminosity and spectra. This involves modelling the spectro-photometric properties of their stellar populations, as well as making corrections for the effects of dust obscuration and ionised gas within each galaxy.

2.10.1 Stellar population synthesis

Given the past star formation history of a galaxy, as well as a choice of IMF, one can calculate the intrinsic luminosity of the galaxy as a function of frequency (or wavelength).

This is known as the *spectral energy distribution* (SED) of the galaxy, $L_\nu^{\rm galaxy}$, which can be expressed as the convolution,

$$L_\nu^{\rm galaxy} = \int_0^t \int_0^\infty \dot{M}_\star \left(t', Z'\right) L_\nu^{\rm SSP} \left(t - t', Z', \phi\right) {\rm d}t' {\rm d}Z', \tag{2.69}$$

where $L_\nu^{\rm SSP}$ is the SED of each stellar population in the galaxy, which is given by,

$$L_\nu^{\rm SSP}(t, Z, \phi) = \int_{M_{\rm min}}^{M_{\rm max}} \phi\left(M'\right) L_\nu^{\rm star}(t, Z) {\rm d}M', \tag{2.70}$$

where $M_{\rm min}$ and $M_{\rm max}$ are the minimum and maximum stellar masses described by the IMF, $\phi(M)$, and $L_\nu^{\rm star}$ is the SED of a single star formed with metallicity, Z, and with an age, t. Libraries of SEDs for stellar populations with different ages, metallicities and choice of IMF have been computed by several authors (e.g. Bruzual A. & Charlot, 1993; Bruzual & Charlot, 2003; Maraston, 2005; Dotter et al., 2007; Lee et al., 2009), based upon theoretical models of stellar evolution and stellar atmospheres, as well as observations of stars with known ages and metallicities. By default, GALFORM uses the stellar population SEDs provided by Bruzual & Charlot (2003) and linearly interpolates over this library in both time and logarithm of the metallicity.

Note that the IMF assumed in GALFORM includes a contribution from brown dwarf stars (with mass $M < 0.1\,{\rm M}_\odot$) which contribute mass, but no light, to the stellar population. The fraction of brown dwarfs is specified through the parameter,

$$\Upsilon = \frac{\text{(mass in visible stars + brown dwarfs)}}{\text{(mass in visible stars)}} \tag{2.71}$$

where typically $1 < \Upsilon \lesssim 2$. The effect of including brown dwarfs is to reduce the luminosities of all stellar populations by a factor $1/\Upsilon$.

Given the frequency-dependent filter response for a photometric band, GALFORM is able to use the emitted luminosity of the galaxy to calculate the absolute magnitude of the galaxy (using either the AB or Vega system) as observed in that band. Magnitudes can be calculated in either the rest frame of the galaxy or the observer frame. To calculate magnitudes in the observer frame a frequency shift is first applied to the galaxy SED.

2.10.2 Dust extinction

The presence of dust grains in the interstellar medium and the intergalactic medium can lead to significant attenuation of the optical and ultra-violet (UV) luminosities of a galaxy, since the dust grains both absorb and scatter light at these frequencies. The UV luminosity of a galaxy is often used as an observational indicator for star formation and

it is therefore essential that extinction due to dust is included in the GALFORM model. In general, dust extinction is incorporated by assuming the dust (whose mass is predicted by the chemical evolution model) is mixed together with the stars in the disk of the galaxy in two phases: in clouds and in a diffuse component (see Granato et al. 2000). Assuming a distribution of dust grain sizes, and combining this with the predicted scalelengths of the disk and bulge, allows one to calculate the optical depth and apply the appropriate attenuation to the luminosity at various wavelengths. An alternative to this internal calculation is to combine GALFORM with the predictions of a specro-photometric radiative transfer code, GRASIL (Silva et al., 1998).

Extinction due to dust was introduced to GALFORM by Cole et al. (2000), who parameterised the total dust content using the central, V-band optical depth, $\tau_{\rm V0}$, defined as the extinction optical depth obtained when looking vertically through the galactic disk at the centre of the galaxy (i.e. at $r = 0$). The dust is assumed to be distributed in the galactic disk with an exponential profile in both the radial and vertical directions. Whilst the stars and dust are assumed to have equal radial scalelengths, it is assumed that they have different vertical scale heights. The properties of the dust are chosen to match the extinction law and dust albedo for the Milky Way. Cole et al. assume that $\tau_{\rm V0}$ scales with the dust mass per unit area, which itself scales with the total mass of metals per unit area in the cold gas content, i.e.

$$\tau_{\rm V0} \propto \frac{M_{\rm dust}}{r_{\rm disk}^2} \propto \frac{M_{\rm cold} Z_{\rm cold}}{r_{\rm disk}^2}. \tag{2.72}$$

The relation is normalised by assuming that gas with solar metallicity, $Z = 0.02$, has a dust-to-gas ratio consistent with that of the local interstellar medium, as measured by Savage & Mathis (1979). The central optical depth is then given by,

$$\tau_{\rm V0} = 0.043 \left[\frac{M_{\rm cold} / \left(2\pi h_{\rm R}^2\right)}{\rm M_\odot \, pc^{-2}} \right] \left(\frac{Z_{\rm cold}}{0.02} \right), \tag{2.73}$$

where $h_{\rm R}$ is the radial scalelength of the dust. Taking the value for $\tau_{\rm V0}$ and assuming a random inclination, GALFORM interpolates over the attenuation tables of Ferrara et al. (1999) to determine the attenuation due to the separate disk and bulge components of each galaxy in each photometric band of interest.

A more sophisticated dust calculation can be performed using the spectro-photometric code GRASIL (Silva et al., 1998), which is capable of computing the emission of the stellar population of a galaxy, the absorption and emission of radiation due to dust, as well as radio emission from massive stars (Bressan et al., 2002). The code computes the

attenuation due to dust using radiative transfer of light through a galaxy with an idealised geometry consisting of a disk and a bulge, each of which may contain diffuse and clumpy gas and dust. GRASIL is therefore an extremely valuable tool for studying galaxies at infrared and submillimetre wavelengths and, as such, has been combined with GALFORM on several occasions (Granato et al., 2000; Baugh et al., 2005; Lacey et al., 2008). The major drawback of the GRASIL code is that it can take several minutes to calculate the dust properties for a single galaxy meaning that to run the code on every galaxy in the large population output by GALFORM would be computationally infeasible. A work-around to this problem is to run GRASIL on a small subsample of a few thousand GALFORM galaxies covering a range of different masses, types and evolutionary stages. For wavelength ranges where the galaxy SED is sufficiently smooth a shortcut was devised to fit the output of GRASIL dust emission to minimise the computational cost of running the code for all wavelengths.

2.10.3 Emission line modelling

GALFORM is able to model the emission lines produced due to the recombination of ionised gas in the interstellar medium. The first step is to determine the number of Lyman continuum photons that are produced, typically by young, massive stars. The production rate of these photons is,

$$\dot{n}_{\rm Lyc} = \int_{\nu_0}^{\infty} \frac{L_\nu^{\rm galaxy}}{h\nu} {\rm d}\nu \tag{2.74}$$

where $\nu_0 = 2.73 \times 10^{21}\,{\rm Hz}$ is the Lyman limit frequency (i.e. the minimum frequency necessary for a photon to have sufficient energy to ionise a hydrogen atom). These photons are assumed to be absorbed by hydrogen atoms in the interstellar medium, thus ionising these atoms. However, rapid recombination leads to a cascade of downward radiative transitions, each emitting a photon, until the atom has returned to the ground state. Through modelling of the number of transitions and the deviation from dynamic equilibrium, GALFORM is able to predict the luminosities and equivalent widths of various emission lines, including, for example, the Hα line.

2.10.4 IGM absorption

Neutral hydrogen in the intergalactic medium that lies along the line-of-sight between a galaxy and an observer will absorb photons in the Lyman series and continuum and so modify the SED of the galaxy. This will, for instance, affect the UV luminosity of a galaxy

and so may further affect estimates of star formation rates in galaxies at high redshift. To model the effects of IGM absorption, estimates of the optical depth of the IGM are required. For this calculation, GALFORM is able to adopt the optical depth models of either Madau (1995) or Meiksin (2006).

2.11 Observational constraints

In order to constrain the free parameters used in GALFORM, the model predictions are compared to several observational statistics, typically of galaxies in the local Universe.

- The galaxy luminosity function is one of the most important constraints for the GALFORM model. It is one of the most fundamental properties of the galaxy population and also one of the best measured. As a result, an important prerequisite for any realistic GALFORM model is that it be able to match the luminosity function of galaxies in the local Universe, typically in the b_J and K bands. For example, the strength of SNe will affect the rate of star formation in low mass halos, thus modifying the faint-end slope of the luminosity function, whereas the strength of AGN will affect the bright-end break as cooling flows are quenched in the most luminous galaxies hosted by massive halos.

- The Tully-Fisher relation (Tully & Fisher, 1977) is an observationally measured correlation between the luminosity of a galaxy and its maximum circular velocity and forms one of the rungs on the 'cosmological distance ladder'. GALFORM models are constrained by their ability to reproduce the slope and zero-point of this relation. For example, increasing the value of the SNe feedback parameter, $\alpha_{\rm hot}$ (see Section 2.6.1), makes it more difficult to form stars in galaxies with low circular velocities and so the luminosity of these galaxies is reduced, causing the model predictions to bend away from the observed correlation at faint magnitudes.

- Being able to reproduce the distribution of the sizes of galaxy disks is a useful constraint to test whether a GALFORM model is successfully modelling the structural properties and internal kinematics of galaxies. In GALFORM, disk sizes are fundamentally determined by the conservation of the angular momentum gained by the cold gas through tidal torques. Correctly predicting the sizes of galaxy disks therefore helps ascertain whether our treatment of the angular momentum of the cold gas is correct. However, the efficiency of stellar feedback mechanisms has also been

noted to have an effect. For example, stellar feedback that is too weak will lead to efficient star formation in small disks at high redshift. Early attempts to model the formation of galaxy disks, found that gas cooling too quickly leads to the baryons losing most of their angular momentum to the dark matter, leading to galactic disks that were too compact compared to observations. This was described as the *angular momentum catastrophe* (Navarro & White, 1993, 1994). One solution to the problem is to strengthen the feedback. Delayed cooling then reduces the loss of angular momentum from the baryons to the dark matter and leads to more extended galaxy disks.

Observationally, knowledge of any two of galaxy scale size, surface brightness and luminosity allows one to construct a bivariate distribution, with which one can predict the third. Galaxy luminosities are typically measured through a summation of the flux within some aperture placed on the galaxy on the sky. Attempting to repeat this procedure with simulated galaxies that are too large or too small will lead to discrepancies in measured luminosities and, therefore, discrepancies in the estimates for the luminosity function or the scaling relations for these galaxies (e.g. González et al., 2009).

- Comparison of model predictions with the relative abundance of galaxies with different morphological types can help place constraints on the frequency and treatment of galaxy mergers and disk instabilities in the model (since these are the only mechanisms in the model capable of forming elliptical galaxies). Additionally, the strength of feedback mechanisms will affect the measured morphological mix (e.g. González et al., 2009). As described above, changing the strength of feedback will affect the growth rate of galaxy disks, thus changing the number of major mergers or disk instabilities that will occur over a given time interval.

- The cold gas content of spiral galaxies provides a fundamental test for the treatment and efficiency of star formation in GALFORM. Comparisons of the model predictions with the cold gas mass-to-light ratios of galaxies, as a function of galaxy luminosity, provide constraints on the star formation timescale and how this varies for galaxies with different disk circular velocities. Indeed, the sensitivity of the cold mass content to the choice of star formation law allowed Lagos et al. (2011b) to use comparisons with the observed cold gas mass function and global cold gas density to compare the predictions of GALFORM when different star formation laws were incorporated into

the model.

- The mass-metallicity or luminosity-metallicity relation (a correlation between the mass or luminosity of a galaxy and its metallicity) provides a means of constraining the treatment of chemical evolution of galaxies within GALFORM. Since metals in the interstellar medium originate from SNe explosions, we would expect the slope of the luminosity-metallicity relation to show some dependence upon the strength of SNe feedback. Consideration of this led to Cole et al. (2000) introducing an effective yield, $p_{\rm eff} = (1-e)p/(1-R+\beta)$, that is not only dependent upon the yield, p, the recycled fraction, R (as predicted by integrating over the IMF) and the fraction, e, of newly produced metals that are ejected directly into the hot gas phase, but also on the strength of SNe feedback, β. Therefore, since SNe feedback is weaker (smaller β) in more massive galaxies, then more luminous galaxies in deeper potential wells will have a larger effective yield and, therefore, more metal-rich stellar populations.

- Although accurate dynamical measures of the mass-to-light ratios of stellar populations in galaxies are difficult due to difficulties in separating contributions to the mass from stars and from dark matter, the GALFORM predictions of the mass-to-light ratios provide a useful consistency check for parameters such as the fraction of brown dwarfs.

- Comparing the predicted distribution of galaxy colours with the observed distribution provides an additional check for any GALFORM model. The complication with using galaxy colours however is the vast number of dependencies; the colours of galaxies can depend upon the efficiency of star formation (which will depend upon the strength of SNe and AGN feedback), the metallicity of the stars (which will depend on the recycled fraction and the yield which, as we have seen, will depend upon the strength of SNe feedback) as well as the choice of IMF. Font et al. (2008) examined the effect on the colours of satellite galaxies by including a treatment of ram-pressure stripping in a version of the GALFORM model. However, the ability to reproduce global trends in galaxy colours, such as a bimodality of the colour distribution, has helped drive improvements to the GALFORM model (e.g. Bower et al., 2006; González et al., 2009).

Chapter 3

Constructing lightcone mock catalogues

In this chapter we outline the method adopted to construct lightcone mock catalogues. Our scheme shares many features in common with the methods used by Blaizot et al. (2005) and Kitzbichler & White (2007), with some improvements. We begin, however, by providing an overview of the basic procedure for constructing mock catalogues and set out the advantages of using a semi-analytical model of galaxy formation.

3.1 Introduction

A very basic mock catalogue could be constructed by randomly sampling one of the measured statistical distributions that describe the galaxy population (e.g. the luminosity function or stellar mass function). Although the resulting catalogue of galaxies would match that particular statistic (by construction), without any further information about the galaxies, such as their colour or spatial distribution, the mock would be very limited.

Building a more realistic mock catalogue, with positional information and the inclusion of other galaxy properties and their evolution, requires the use of a numerical simulation which follows the growth of structure in the dark matter. The procedure for constructing mock catalogues from a numerical simulation can be broken down into the following steps: (i) generate a population of galaxies either empirically or using a physical model, using either the dark matter distribution or dark matter halos, (ii) place these galaxies into a cosmological volume, (iii) apply the angular and radial selection functions of the survey.

3.1.1 Generating a galaxy population

To generate a population of galaxies one must first model the distribution of dark matter, which is often done with a N-body simulation. Dark matter only N-body simulations allow one to build halo populations using gravity alone. The full spatial information provided by N-body simulations allows one to extract clustering information, which would otherwise

not be available if a Monte-Carlo approach was to be used. Additionally, the merger histories of halos in N-body simulations will also include environmental effects, such as halo assembly bias (Gao et al., 2005).

The way in which dark matter halos are populated with galaxies is where the methods of mock catalogue construction can differ. Blaizot et al. (2005) (see also Baugh, 2008) summarise several of the different methods available, which include using phenomenological models to assign galaxies to dark matter particles in the simulation (e.g. Cole et al., 1998) or using empirically derived statistics, such as the *halo occupation distribution* (HOD, Berlind & Weinberg, 2002; Song et al., 2012) or *sub-halo abundance matching* (Vale & Ostriker, 2004). Other more physical approaches are also possible. For instance, one could include the baryons in the original simulation, using either a grid-based or particle-based method to solve the hydrodynamical equations. The problem with direct, hydrodynamical simulations however, is that they are computationally expensive and so, in practice, are restricted to small volumes (e.g. the $25h^{-1}$Mpc and $100h^{-1}$Mpc boxes used in the Overwhelmingly Large Simulations project of Schaye et al., 2010).

A powerful approach that we choose to adopt is to use a *semi-analytical model* of galaxy formation to populate the halo merger trees extracted from a high resolution, cosmological N-body simulation (Diaferio et al., 1999; Benson et al., 2000a; Blaizot et al., 2005; Kitzbichler & White, 2007; Sousbie et al., 2008; Overzier et al., 2009; Cai et al., 2009; Henriques et al., 2012). Modelling of various physical processes, such as the cooling of gas within dark matter halos, is necessary to follow the baryonic component and predict the fundamental properties of galaxies, such as their stellar mass and star formation history. The adoption of an *initial mass function* (IMF), a *stellar population synthesis* (SPS) model and a treatment of dust extinction allows these fundamental properties to be connected with observables, thus enabling a direct comparison between observations and the predictions of the galaxy formation model.

3.1.2 Generating a cosmological volume

Current and future galaxy surveys are designed to probe ever larger cosmological volumes. As a result there is a growing demand for simulations with boxes of sufficient size to match the volumes of these surveys. Unfortunately, current computing power means that a compromise must often be made between the volume of the simulation box and the resolution at which the simulation is carried out. Therefore a sufficiently large cosmological volume can often only be sampled by tiled replication of a smaller box simulation. For very

shallow galaxy surveys (e.g. with a median redshift $z \lesssim 0.05$), the lookback time is sufficiently small that typical galaxy properties will not have undergone significant evolution across the redshift interval covered by the survey. In these instances, the statistics of the galaxy population at the extremes of the survey will not be too dissimilar to the statistics today and so one could build a mock catalogue using galaxies from a single simulation snapshot. However, for very deep galaxy surveys, which cover a significant lookback time, we would expect to see substantial evolution in galaxy properties and in the growth of large-scale structure. Therefore, more sophisticated mock catalogues that tile the survey volume using many different simulation snapshots are required to adequately reproduce the evolution seen in the properties of galaxies and their clustering. The mock catalogues that we construct in this work are *lightcone* mock catalogues, in which galaxies are placed according to the epoch at which they first cross the observer's past lightcone, i.e. at the location at which the light emitted from the galaxy has just enough time to reach the observer, and thus incorporate the evolution of structure with cosmic time.

Finally, observations will be subject to uncertainties or biases, introduced as a result of survey design or selection effects, and so to properly relate theoretical predictions to observations we must subject the simulated data to the same selection functions as the observed galaxy sample.

3.1.3 Why use a semi-analytical galaxy formation model?

Modelling the formation of galaxies is a difficult task. Part of the problem is that our knowledge of the underlying physics is limited and so we cannot simply write down a precise formulation for every process. Furthermore, despite the continued development of direct, hydrodynamic simulations, current computational capabilities mean that many of the relevant processes (for example, star formation or supernova feedback) remain firmly below the resolution limits of direct simulations and can only be addressed through "sub-grid" physics. Semi-analytic models describe the sub-grid physics using physically motivated, parameterised equations that follow the evolution of baryons trapped in the gravitational potential wells of hierarchically grown dark matter halos (White & Rees 1978; Cole 1991; White & Frenk 1991; Kauffmann et al. 1993; Cole et al. 1994; for reviews of the semi-analytical approach see Baugh 2006 and Benson 2010). There are several compelling advantages to using semi-analytic models for building mock catalogues:

1. The development of deep, wide-field photometric galaxy surveys spanning large

cosmological volumes has led to demand for large (suites of) mock catalogues that can be constructed rapidly and accurately. Semi-analytic modelling is currently the only physical approach that meets these ideals: such models are capable of populating large cosmological volumes with galaxies much faster and at a lower computational cost than is currently possible with hydrodynamical simulations.

2. The modular design of semi-analytic models allows new physics to be incorporated readily. Combined with their short run-time, this means that semi-analytic models can be tuned to match observations quickly, in response to a change to the background cosmology or to the galaxy formation physics. Moreover, the larger computational box that can be used in the N-body and semi-analytical approach, compared with a hydrodynamical simulation, means that the clustering predictions are robust out to larger scales.

3. Empirically motivated methods, such as HOD modelling, must first be calibrated against observational data, (e.g. the LasDamas mock catalogues, McBride et al., 2009). Hence, mock catalogues built using such methods are limited by the availability of observational data at high redshift. Furthermore, the data that is available may be affected by sample variance leading to unrepresentative HOD parameters being fitted. Semi-analytic models, however, once tuned to fit the observations of galaxies at low redshift, can *predict* galaxy properties out to high redshift, without further observational input. Mock catalogues built from semi-analytical models are therefore much more flexible than catalogues constructed using other methods.

4. The next generation of galaxy surveys will map the sky across a large portion of the electromagnetic spectrum, with multi-wavelength follow-up observations resulting in potentially complex survey selection functions, such as for the Galaxy And Mass Assembly (GAMA) Survey (Driver et al., 2011). Ideally mock catalogues for future surveys need to provide a diverse range of galaxy properties as well as providing the capability to select galaxies simultaneously in multiple bands. Semi-analytic models model the complete star formation history for each galaxy and so can predict many different galaxy properties. Mock catalogues based upon semi-analytic models are already capable of mimicking sophisticated multi-band selection criteria.

3.2 Constructing a galaxy lightcone

By first running the GALFORM model[1] on the halo merger trees of the *Millennium Simulation* we generate a galaxy population that is used to build the lightcone catalogues. Galaxy properties are stored for each fixed, snapshot epoch that falls within the redshift range of interest for a particular survey.

An observer is then placed inside the simulation box at a position that can be set manually[2] or at random.

3.2.1 Replication of the simulation box

The cosmology used in the Millennium Simulation means that the simulation box side-length, $L_{\rm box} = 500h^{-1}$Mpc, corresponds to the co-moving distance out to $z = 0.17$. Therefore, in order to generate a cosmological volume that is of sufficient size to fully contain any galaxy survey that extends out to a modest redshift, it is necessary to tile replications of the simulation box (see the discussion in Section 3.1.2).

The number of replications per axis, $n_{\rm rep}$, that need to be stacked around the original box (containing the observer) is given by[3],

$$n_{\rm rep} = \left\lfloor \frac{r_{\rm max}}{L_{\rm box}} \right\rfloor + 1, \tag{3.1}$$

where $r_{\rm max}$ is the maximum co-moving radial distance that we want to reach in the final mock catalogue. Including the original simulation box, we have a total of $(2n_{\rm rep}+1)^3$ replications. The Cartesian co-ordinate system, $(\hat{\mathbf{X}}, \hat{\mathbf{Y}}, \hat{\mathbf{Z}})$, of the combined 'super-cube' is then translated so that the observer is located at the origin.

An unfortunate consequence of generating a large volume in this way is that structures can appear repeated within the final lightcone volume. Although repeated structures cannot have a co-moving separation less than the simulation box side-length, if any repeated structures have small angular separations when projected onto the 'mock sky', then projection-effect artefacts can be introduced into the catalogue. Blaizot et al. (2005) illustrate the effect of these artefacts, along with possible methods for eliminating them. One method that they demonstrate to be effective is to apply random sequences of $\pi/2$

[1] We stress that our lightcone construction algorithms are independent of choice of semi-analytic model and can be run using any input galaxy formation model.

[2] Often one will choose to position the observer manually if they desire the observer to be placed in a specific location, such as an environment similar to the Local Group.

[3] $\lfloor x \rceil$ means that x is rounded down to the nearest integer.

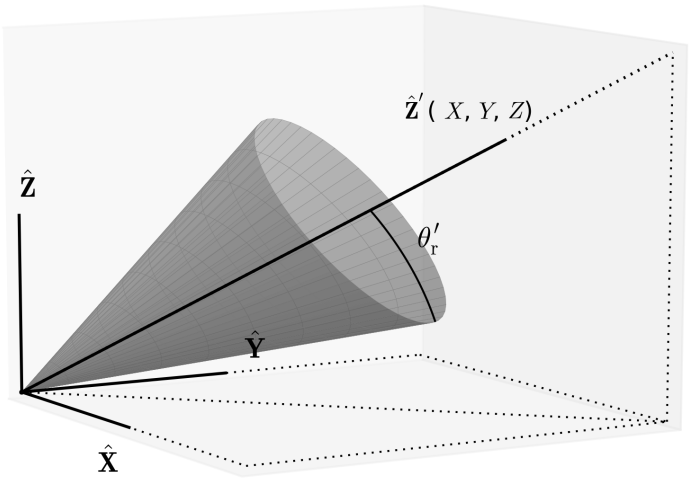

Figure 3.1: Schematic of lightcone geometry. The axis $\hat{\mathbf{Z}}'$ defines the central *line-of-sight* vector of the observer. The angle θ'_r defines the angular size of the field-of-view of the lightcone. Any galaxy whose position vector, $\tilde{\mathbf{r}}'(X', Y', Z')$, is offset from the $\hat{\mathbf{Z}}'$ axis by an angle $\theta' > \theta'_r$ is excluded from the lightcone.

rotations and reflections to the replicated boxes so that any repeated structures are viewed at different orientations and appear as different structures. The problem with this approach is that, due to the periodic boundary conditions of the N-body simulation, when tiling the replications, any transformation besides a translation would add undesirable discontinuities into the underlying density field. However, if, for example, one wishes to extract clustering statistics, the underlying density field should be preserved. We therefore choose not to use this method in the construction of our lightcone catalogues.

3.2.2 Orientating the observer

Our aim when orientating the observer is to be able to define a right-handed Cartesian co-ordinate system, $(\hat{\mathbf{X}}', \hat{\mathbf{Y}}', \hat{\mathbf{Z}}')$, such that the observer is looking down the $\hat{\mathbf{Z}}'$ axis, as illustrated in Figure 3.1. This axis defines the central axis of symmetry of the conical volume of the lightcone and points to the centre of the field of the lightcone on the mock sky, i.e. $\hat{\mathbf{Z}}'$ points along the central *line-of-sight* vector of the observer. The half-opening angle, θ'_r, governs the angular extent of the field-of-view of the lightcone (see Section 3.2.3). The orientation of the observer is simply how we describe this vector, $\hat{\mathbf{Z}}'$, in terms of the global Cartesian axes of the "super cube", $\hat{\mathbf{Z}}'(X, Y, Z)$.

For deep, pencil-beam mock catalogues, carefully choosing the orientation of the observer can minimise, or even remove, structure repetition. The approach adopted by Kitzbichler & White (2007) is to orientate the observer in a 'slanted' direction, with respect to the Cartesian axes of the simulation box, so that the observer is not looking along any of the Cartesian axes or the cube diagonals (along which structure repetition can introduce noticeable artefacts). By defining the central line-of-sight of the observer as $\hat{\mathbf{Z}}'(X, Y, Z) = (n, m, nm)$, where m and n are integers with no common factor, Kitzbichler & White are able to construct lightcone catalogues with a near-rectangular sky coverage of $1/m^2n \times 1/n^2m$ (radians) in which the first repeated structure will lie at distance of $\sim mnL_{\rm box}$ from the observer.[4] When constructing lightcone catalogues for which we wish to minimise duplicated structures, albeit at the expense of the solid angle of the catalogue, we adopt this approach. This is necessary for applications considering, for example, the angular clustering of galaxies, where projection effects could severely distort the clustering signal.

Once we are satisfied with the chosen orientation of $\hat{\mathbf{Z}}'$, we define the axis $\hat{\mathbf{X}}'$ (to be perpendicular to both $\hat{\mathbf{Z}}'$ and $\hat{\mathbf{X}}$) and the axis $\hat{\mathbf{Y}}'$ (to be perpendicular to both $\hat{\mathbf{X}}'$ and $\hat{\mathbf{Z}}'$).

3.2.3 Finalising the lightcone geometry

Now that we know the location and orientation of the observer we can set about applying the necessary geometrical cuts to construct the lightcone volume. The first step is to isolate a spherical volume about the observer, with a co-moving radius, $r_{\rm max}$, whose value is sufficiently large that, given the flux limits of the survey we wish to emulate, we would expect to be well into the high-redshift tail of the galaxy redshift distribution (such that only a negligible fraction of the brightest, high redshift galaxies are missed). This radial cut is applied to help speed up the calculation so that we are not searching for galaxies in box replications that are too far from the observer to contribute a significant number of objects to the mock catalogue. For boxes with a fraction of their volume lying within $r_{\rm max}$, we check when each galaxy will enter the observer's past lightcone (see Section 3.2.4). If a galaxy enters the lightcone at a distance greater than $r_{\rm max}$ (or it never enters the lightcone at all) then it is discarded.

[4] Carlson & White (2010) adopt a similar approach to Kitzbichler & White by performing volume remapping of the original simulation box such that the mock catalogue geometry can fit inside the new geometry without the need for box replication.

Next we apply an angular cut on the mock galaxies, which is dictated by the solid angle of the galaxy survey we wish to mimic. The solid angle in steradians, Ω, of the mock catalogue is defined by

$$\Omega = 2\pi \left[1 - \cos\left(\theta'_r\right)\right], \tag{3.2}$$

where θ'_r is the field-of-view angle of the catalogue. By varying the value of θ'_r we can construct lightcones with solid angles ranging from pencil beams, to all-sky (4π) surveys.[5] Following this cut, the catalogue volume resembles the sector of a sphere, with half-opening angle θ'_r and $\hat{\mathbf{Z}}'$ as its axis of symmetry. For those boxes whose volume overlaps that of the catalogue, we calculate the position at which each galaxy enters the lightcone. Using this position we calculate the angle θ', the angle between the position vector of the galaxy and the $\hat{\mathbf{Z}}'$ axis, and discard any galaxy with $\theta' > \theta'_r$, that lies outside the solid angle of the catalogue.

Finally, for those galaxies that are successfully included in the lightcone, we determine their right-ascension, α, and declination, δ, on the mock 'sky'. We do this by first defining a sky coordinate system such that the observer's central line-of-sight vector, $\hat{\mathbf{Z}}'$, points towards a right ascension, α_0, and declination, δ_0, on the sky. We then determine the sky position of a galaxy by passing $\tilde{\mathbf{r}}(X, Y, Z)$, through the transformation, $\mathcal{R}_Z(\alpha_0)\mathcal{R}_Y(\pi/2 - \delta_0)\ \tilde{\mathbf{r}}$, where $\mathcal{R}_Z$ and $\mathcal{R}_Y$ are the standard 3-dimensional Cartesian rotation matrices about the $\hat{\mathbf{Z}}$ and $\hat{\mathbf{Y}}$-axes respectively. (We assume that lines of constant declination lie parallel to the $X - Y$ plane, so do not apply any rotation about the $\hat{\mathbf{X}}$ axis).

3.2.4 Positioning galaxies within the lightcone

The lightcone selection of galaxies is carried out by identifying those galaxies whose light has sufficient time to reach the observer. However, before one can calculate when a galaxy enters the lightcone, one must determine the epoch at which its host dark matter halo enters the observer's past lightcone.

[5] By setting $\theta'_r = \pi$ we can construct all-sky lightcone catalogues. When constructing such catalogues we can apply an additional geometrical cut to remove galaxies that would be obscured by the plane of the Milky Way. Having calculated the celestial co-ordinates of the galaxy on the mock sky, we determine the galactic latitude, b, of the galaxy and reject all galaxies with $|b| < b_{\rm lim}$, where $b_{\rm lim}$ is the user-specified galactic latitude limit. The solid angle of the all-sky lightcone is then calculated as, $\Omega_{\rm all-sky} = 4\pi - 2\pi\left[\sin\left(b_{\rm lim}\right) - \sin\left(-b_{\rm lim}\right)\right]$.

Placement of halo centres

A halo, located at $\tilde{\mathbf{r}}(X, Y, Z, t)$, at a lookback time, t, will be "visible" to the observer at all instances where $|\tilde{\mathbf{r}}(X, Y, Z, t)| \leqslant r_{\rm c}(t)$, where $r_{\rm c}$ is the maximum distance that light could have travelled in the time t, i.e. the maximum co-moving, radial distance that is visible to the observer. For a flat cosmology, i.e. with $\Omega_k = 0$, at the epoch corresponding to redshift, z, the maximum co-moving, radial distance, $r_{\rm c}$, that is visible to an observer at the present day is given by,

$$r_{\rm c}(z) = \int_0^z \frac{c\,{\rm d}z'}{H_0\sqrt{\Omega_{\rm M}\,(1+z')^3 + \Omega_\Lambda}}, \tag{3.3}$$

where H_0 is the Hubble Constant at the present day, $\Omega_{\rm M}$ is the matter density of the Universe and Ω_Λ is the vacuum energy density of the Universe but at the present day. To construct a mock galaxy catalogue, we place each halo at the epoch at which it enters the observer's past lightcone, i.e. the epoch at which the halo would first become "visible" to the observer. If this epoch corresponds to a redshift, z, then the halo is placed at the position, $\tilde{\mathbf{r}}(X, Y, Z, z)$, at which

$$|\tilde{\mathbf{r}}(X, Y, Z, z)| = r_{\rm c}(z). \tag{3.4}$$

Each snapshot, i, in the *Millennium Simulation* corresponds to a discrete cosmic epoch, with redshift, z_i. To determine when a halo enters the lightcone, we loop over the simulation snapshots searching for the time-step during which Eq. (3.4) is satisfied. By comparing the position, $\tilde{\mathbf{r}}_j(X_j, Y_j, Z_j, z_i)$, of a halo j that exists at z_i to the maximum co-moving distance, $r_{\rm c}(z_i)$, that is visible at that epoch, and doing the same for the descendant of the halo, labelled k, that exists at the next snapshot, $z_{i+1} < z_i$, we can determine whether halo j will enter the lightcone between the snapshots i and $i+1$, i.e. whether $z_{i+1} < z < z_i$.

Using $\tilde{\mathbf{r}}_j(X_j, Y_j, Z_j, z_i)$ and $\tilde{\mathbf{r}}_k(X_k, Y_k, Z_k, z_{i+1})$ as boundary conditions, we interpolate along the orbital path of the halo j to find the exact epoch, z, at which it enters the lightcone and the position, $\tilde{\mathbf{r}}_j(X, Y, Z, z)$, at which this occurs. We use a cubic polynomial to describe the position of the halo, in each Cartesian direction, as a function of the time, t, between the adjacent snapshots (i.e. $t_{i+1} < t < t_i$). For example, the Cartesian X component of the path is given by,

$$X(t) = A_X t^3 + B_X t^2 + C_X t + D_X, \tag{3.5}$$

where A_X, B_X, C_X and D_X are coefficients that can be determined by requiring that the boundary conditions ($X(t = t_i) = X_j(t_i)$, $X(t = t_{i+1}) = X_k(t_{i+1})$, $\dot{X}(t = t_i) = \dot{X}_j(t_i)$, $\dot{X}(t = t_{i+1}) = \dot{X}_k(t_{i+1})$) are satisfied. The X component of the velocity of the galaxy at time, t, is given by the derivative of Eq.(3.5) with respect to time. Equations similar to Eq.(3.5) can be derived for the $Y(t)$ and $Z(t)$ components. The centre of mass of the halo is then placed at $\tilde{\mathbf{r}}_j(X, Y, Z, z)$.

Our decision to use interpolation to determine halo positions is an extension of the method of Kitzbichler & White (2007), who explicitly chose not to use interpolation but instead placed halos according to the snapshot with the epoch closest to the one at which the halo enters the lightcone. Kitzbichler & White adopted this approach because of the difficulties inherent in using interpolation to predict realistic orbital paths for satellite galaxies. In the next section, we discuss these difficulties and suggest a solution that provides a good approximation for our purposes.

Placement of galaxies

The finite spatial extent of halos means that central and satellite galaxies within a halo will enter the lightcone at slightly different times. Central galaxies are positioned on the most bound particle of the most massive `SUBFIND` group (see Section 2.2.2) and are defined to be at rest relative to the halo. The location and time at which a central galaxy enters the lightcone is thus equal to that of its host halo and so for these galaxies we can use the calculation for the halo centre, as presented in Section 3.2.4. However, satellite galaxies can enter the lightcone at an earlier or later epoch than the centre of the host halo. When positioning a satellite galaxy we can still interpolate over the evolutionary path of its host halo, but we must first correct the spatial positions along the path to account for the relative offset between the position of the satellite galaxy and the centre of the halo. Therefore, we first need a model to describe the orbital path of a satellite within its host halo.

Modelling physically viable satellite orbits is a non trivial task. Difficulties arise when the large orbital velocities of satellite galaxies lead to orbital time-scales that are much shorter than the spacing of the simulation snapshots. Care must therefore be taken to ensure that numerical artefacts do not introduce large positional errors, which might in turn lead to inaccurate predictions for the one-halo term in the galaxy correlation function.

If, for example, we attempt to describe the orbital path of a satellite galaxy using a

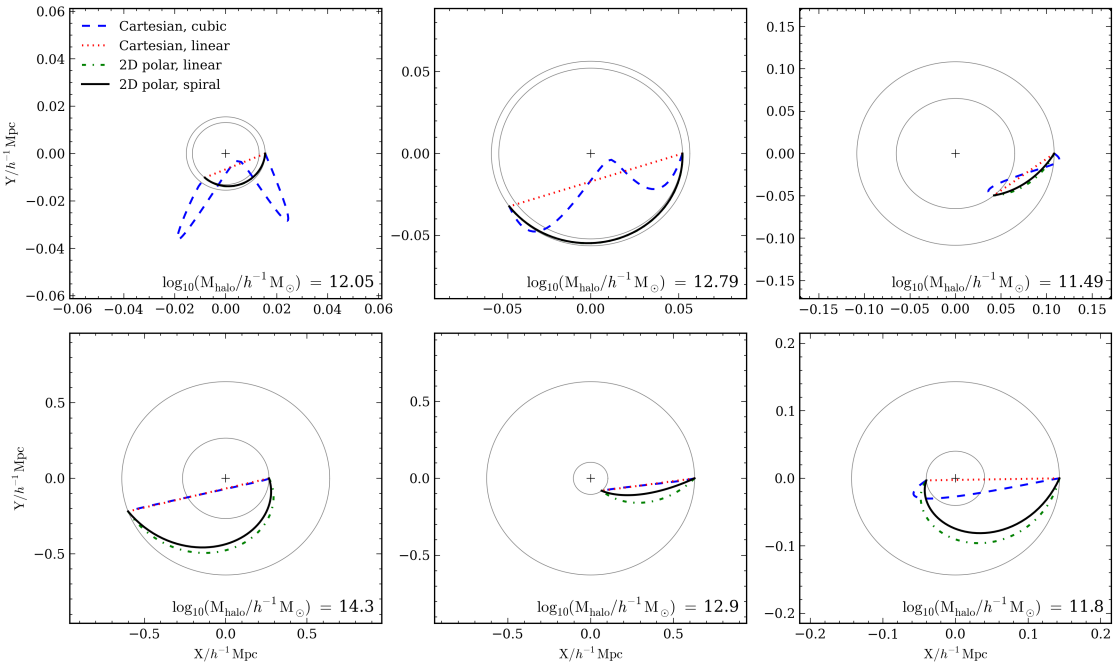

Figure 3.2: Examples demonstrating the modelling of the orbital paths of satellite galaxies between two adjacent simulation snapshots using different interpolation schemes. The positions of the satellite galaxies are displayed relative to the centre of mass of the halo, which is marked with a +. Circles show circular orbits at the start and end radii of the path of the satellite galaxy. The various interpolation schemes, which use either 3-dimensional Cartesian co-ordinates or 2-dimensional polar co-ordinates, are discussed in Section 3.2.4 and are shown using different line colours and styles, as indicated by the key in the top left panel.

cubic polynomial in Cartesian space that is constrained to satisfy both the position and velocity boundary conditions, then in rare instances we may find that the large orbital velocities of satellite galaxies lead to orbital paths that are highly eccentric and extend out to large orbital radii. In the majority of cases where orbital velocities are small, the cubic function fits an orbital path similar to a simple linear interpolation scheme, which ignores the velocity boundary conditions. Example orbits, modelled using different interpolation schemes, are shown in Figure 3.2. Unfortunately, if a halo and its descendent are found on opposing sides of the halo centre of mass, then these interpolation schemes would lead to satellites being positioned much closer to the centre of mass of the halo than they should be. This would have the effect of boosting the clustering signal on small scales, as shown in the two left-hand panels of Figure 3.3.

Since measurements of galaxy clustering are integral for achieving many of the goals set by current and future galaxy surveys, we choose to prioritise the preservation of the galaxy clustering signal in real space. We do this by moving to a 2-dimensional (2D)

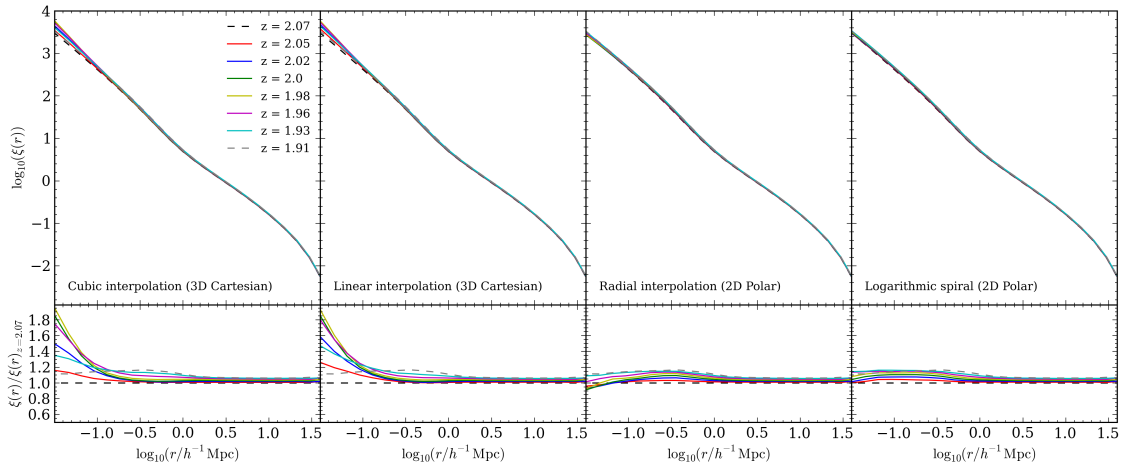

Figure 3.3: The real-space correlation function of galaxies predicted using four different satellite interpolation schemes: cubic interpolation in 3D Cartesian space (far left), linear interpolation in 3D Cartesian space (middle left), radial interpolation in 2D polar space (middle right) and modelling the satellite orbits using a logarithmic spiral in 2D polar space (far right). The upper panels show the correlation function for galaxies at two adjacent simulation snapshots (corresponding to redshifts $z = 1.91$ and $z = 2.07$, grey and black dashed lines) and the same galaxies at six intermediate redshifts (various solid, coloured lines). The lower panels show the ratio of each correlation function, relative to the correlation function measured at the $z = 2.07$ snapshot.

plane, defined by the position of the halo centre of mass and the positions of the satellite, j, and its descendent, k, relative to the halo centre. By assuming that the orbit of satellite j is restricted to this plane, we use linear interpolation to express the change in the 2D polar co-ordinates of the orbit of j (relative to the centre of mass of the halo, located at the origin) as a function of time between the snapshot epochs, $t_{i+1} < t < t_i$.

We describe the angular position, ϕ, of the satellite along its orbit as

$$\phi(t) = \phi_j(t_i) + [\phi_k(t_{i+1}) - \phi_j(t_i)] \left[\frac{t - t_i}{t_{i+1} - t_i} \right]. \tag{3.6}$$

A caveat is that we perform the interpolation along the path that corresponds to the smallest angular separation between the position of a satellite and its descendent, which may lead to satellites changing directions. To describe the change in the radius, ρ, of the orbit of the satellite we can choose to either linearly interpolate the radius in the same way, i.e.

$$\rho(t) = \rho_j(t_i) + [\rho_k(t_{i+1}) - \rho_j(t_i)] \left[\frac{t - t_i}{t_{i+1} - t_i} \right], \tag{3.7}$$

or couple the radius to the angle using a simple model, such as a logarithmic spiral,

$$\rho(t) = a \cdot \mathrm{e}^{b \cdot \phi(t)}, \tag{3.8}$$

where $a = \rho_j(t_i)$ and $b = \phi_k(t_{i+1}) \ln(\rho_k(t_{i+1})/\rho_j(t_i))$. Note that in these two cases we are ignoring the velocity boundary conditions and assuming that $\ddot{\rho}(t) = \ddot{\phi}(t) = 0$. However, as can be seen in Figure 3.3, these methods preserve the galaxy clustering to smaller scales than possible with the 3-dimensional (3D) cubic or linear approaches. Note that in the above cases we can interpolate the orbital velocities of satellites using the same equations as used for the positions.

By converting back to 3D Cartesian coordinates we can express the epoch, z, at which the satellite enters the lightcone as the position at which,

$$|\tilde{\mathbf{r}}_{\rm halo}(X, Y, Z, z) + \tilde{\mathbf{r}}'_j(X, Y, Z, z)| = r_{\rm c}(z), \tag{3.9}$$

where $\tilde{\mathbf{r}}_{\rm halo}(X, Y, Z, z)$ is the global position of the dark matter hosting the satellite at this epoch and $\tilde{\mathbf{r}}'_j(X, Y, Z, z)$ is the position of the satellite relative to the centre of this halo.

3.2.5 Treatment of galaxy properties in the lightcone

Intrinsic properties

For each galaxy that enters the lightcone and satisfies the geometrical cuts described in Section 3.2.3, we need to output galaxy properties (i.e. stellar mass, SFR, etc.) that are appropriate for the epoch at which we have placed the galaxy.

With knowledge of the star-formation history of the galaxy, we can follow the evolution of any galaxy property over cosmic time. However, as with galaxy positions, this information is only recorded at the discrete epochs corresponding to the simulation snapshots. Ideally we would like to again use interpolation to determine the value for any galaxy property at any given epoch. Unfortunately the evolution of the majority of galaxy properties is complex and by interpolating between the snapshot epochs we risk over-simplifying this evolution and deriving incorrect values. For instance, the build-up of the stellar mass of a galaxy between two consecutive snapshots will receive contributions from many different sources. Besides quiescent star formation in the disk, many other events, such as disk instabilities or mergers with one or more other galaxies, can lead to starbursts and a SFR that is highly variable with time. In the case of galaxy mergers we cannot accurately say from the snapshot data alone, when, during a time-step, the merger occurred. Therefore interpolation over the properties of each progenitor may lead to double counting and, at the epochs at which they enter the lightcone, each progenitor having estimated properties that are incorrect (possibly significantly).

We could evaluate galaxy properties for any given epoch by solving the set of coupled differential equations that govern the exchange of material between the hot gas in the halo and the cold gas and stars in the galaxy. However, this exercise is non-trivial and would require the full calculation performed by GALFORM to be reproduced for each galaxy out to the epoch at which it enters the lightcone, which would be computationally expensive. Similarly, we could have originally run GALFORM and output the galaxy properties on a finer time mesh. However, this would extend the run time of the model and take up significantly more disk space. Instead, we adopt a procedure similar to that of Kitzbichler & White (2007) and, for any galaxy that enters the lightcone, we assign the galaxy the intrinsic properties it had at the snapshot immediately prior to the epoch, z, at which it entered the lightcone, i.e. the snapshot, i, with the smallest redshift, z_i, for which $z_i > z$.

Observed properties

Having set the intrinsic properties of the galaxies we can now use this information, along with their positions, to evaluate their observed properties, namely their observed fluxes (and apparent magnitudes). At this point we need to use the position of a galaxy to derive its luminosity distance, d_L, which is required to relate the emitted luminosity per unit frequency, $L_\nu(\nu_\mathrm{e})$, of an object to its observed flux per unit frequency, $S_\nu(\nu_\mathrm{o})$. For a flat Universe the luminosity distance out to a redshift, z, is simply, $d_L(z) = r_c(z)(1+z)$. Therefore a galaxy in the lightcone at a cosmological redshift, z, will have an observed flux,

$$S_\nu(\nu_\mathrm{o}) = (1+z)\,\frac{L_\nu(\nu_\mathrm{e})}{4\pi d_L^2\,(z)}, \tag{3.10}$$

where ν_o is the observed frequency of the light from the galaxy. The emitted (rest-frame) frequency is related to the observed (observer-frame) frequency by $\nu_\mathrm{e} = \nu_\mathrm{o}\,(1+z)$. The observer-frame apparent magnitude of a galaxy, in the AB system, is then given by,

$$m_\mathrm{AB} = -2.5\log_{10}\left[\frac{\int S_\nu(\nu_\mathrm{o})R(\nu_\mathrm{o})\mathrm{d}\nu_\mathrm{o}}{S_{\nu_\mathrm{o}}\int R(\nu_\mathrm{o})\mathrm{d}\nu_\mathrm{o}}\right], \tag{3.11}$$

where $R(\nu_\mathrm{o})$ is the filter response of a specified photometric band and S_{ν_o} is the AB reference flux per unit frequency (Oke & Gunn, 1983).

In our case GALFORM calculates the emitted luminosity of a galaxy, so we can calculate the observer-frame absolute magnitude, M_AB, of the galaxy (assuming $z \neq 0$) from,

$$M_\mathrm{AB} = -2.5\log_{10}\left[\frac{\int L_\nu(\nu_\mathrm{e})R(\frac{\nu_\mathrm{e}}{1+z})\mathrm{d}\nu_\mathrm{e}}{L_{\nu_\mathrm{o}}\int R(\frac{\nu_\mathrm{e}}{1+z})\mathrm{d}\nu_\mathrm{e}}\right], \tag{3.12}$$

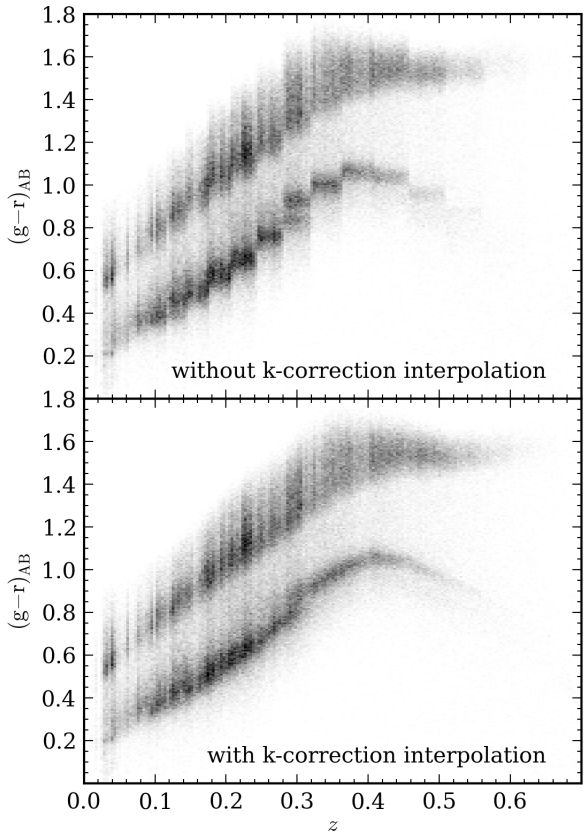

Figure 3.4: Distribution of SDSS $g - r$ colour as a function of redshift, z, in a lightcone catalogue constructed for the GAMA survey, both without k-correction interpolation (upper panel) and with k-correction interpolation (lower panel, see Section 3.2.5 for details). Shading corresponds to the number density of galaxies. (Note that the apparent vertical stripes in the galaxy distributions correspond to peaks in the galaxy redshift distribution.)

where L_{ν_o} is now the AB reference luminosity, $L_{\nu_o} = 4\pi(10\mathrm{pc})^2 S_{\nu_o}$. From this we can calculate the observer-frame apparent magnitude of a galaxy, in the AB system, by,

$$m_{\mathrm{AB}} = M_{\mathrm{AB}} + 5\log_{10}\left(\frac{d_L(z)}{10\mathrm{pc}}\right) - 2.5\log_{10}(1+z)\,. \tag{3.13}$$

Due to the large number of galaxies modelled by `GALFORM`, the full SED of each galaxy is not stored. Instead, the luminosity is computed in a set of filters specified at run time. Hence the definition of the filter response in the galaxy rest frame, $R(\nu_e/(1+z))$, is tied to the output redshifts of the simulation snapshots and the k-correction applied does not correspond to the redshift of the galaxy in the lightcone. This discrepancy leads to visible discontinuities in distributions involving the photometric properties of the galaxies, such as galaxy colours versus redshift, as shown in the upper panel of Figure 3.4. The breaks apparent in the distribution correspond to the redshifts of the simulation snapshots.

As discussed in the previous section, the complex time dependence of galaxy luminosity means that we cannot simply interpolate the absolute magnitudes. However, since the size of the wavelength shift applied to a filter depends only on the redshift to the galaxy (and not on any of its intrinsic properties), then we can apply a correction to all the observer-frame absolute magnitudes (and dust emission luminosities) to take into account the redshift of lightcone crossing. Consider again a galaxy, j, that is originally found in the snapshot, i, at redshift z_i, and which has a descendent in snapshot $i+1$, at a redshift $z_{i+1} < z_i$. Assume that the galaxy has an observer-frame absolute magnitude $M_j(z_i)$. Since the wavelength shift applied depends only on the redshift of a galaxy, we can easily predict the observer-frame absolute magnitude that j would have if placed at the redshift of its descendent in snapshot $i+1$, i.e. $M_j(z_{i+1})$ within the `GALFORM` code, but with a star-formation history computed up to z_i. If the galaxy, j, enters the lightcone at an intermediate epoch, z, then we can interpolate between $M_j(z_i)$ and $M_j(z_{i+1})$ to estimate $M_j(z)$. Note that by interpolating the magnitudes (and luminosities) in this way, we have not changed the shape of the SED of the galaxy, but rather have applied a further systematic wavelength shift to the galaxy SED. As can be seen in the lower panel of Figure 3.4, this correction, which was also applied by Blaizot et al. (2005) and Kitzbichler & White (2007), smooths out the 'saw-tooth' pattern seen in the colour distribution.

We can also calculate an observed redshift for the mock galaxies, emulating the measurement that would be taken from a galaxy spectrum using one or more identified emission lines. The observed redshift, z_{obs}, of a galaxy, which includes the cosmological redshift due to the Hubble flow as well as a component due to the local peculiar motion of the

galaxy, is defined by,

$$z_{\rm obs} = (1+z)\left(1+\frac{v_{\rm r}}{c}\right) - 1, \tag{3.14}$$

where z is the cosmological redshift at which the galaxy enters the lightcone and $v_{\rm r}$ is the radial component of the peculiar velocity vector, $\tilde{\mathbf{v}}$, of the galaxy (i.e. $v_{\rm r} = \tilde{\mathbf{v}} \cdot \hat{\mathbf{r}}$, where $\hat{\mathbf{r}}$ is the normalised line-of-sight position vector of the galaxy). We do not at present include any calculation of photometric redshifts, or their uncertainties, in `GALFORM` or our lightcone code. These properties can be readily calculated in post-processing using the photometry output for each galaxy.

3.2.6 Applying the survey criteria

The final stage in constructing a mock catalogue is to apply the radial selection criteria of the survey being mimicked and reject those galaxies fainter than the specified flux limits. For many surveys this involves placing a cut on the flux at one or more wavelengths or an apparent magnitude limit in one or more photometric bands. We can select galaxies according to any intrinsic or observed galaxy property. For example, besides generating flux limited lightcone catalogues, we are able to construct catalogues limited by stellar mass, atomic hydrogen mass or even halo mass. Given a list of selection criteria, we can control whether a galaxy must pass just one or all of these criteria simultaneously in order to be included in the final catalogue. Note that the lightcone catalogues that we produce correspond to ideal surveys, i.e. we do not apply any completeness masks or simulate the loss of galaxies due to poor observing conditions, fibre collisions or quality of spectra. Such completeness effects are survey specific and can be applied to the catalogues in post-processing.

3.3 Conclusions

We have presented a method for constructing end-to-end mock galaxy catalogues by applying a semi-analytical model of galaxy formation to the halo merger trees extracted from a cosmological N-body simulation. The mocks that we construct are *lightcone* catalogues, in which a galaxy is placed according to the epoch at which it first enters the past lightcone of the observer. Thus our catalogues incorporate the evolution of galaxy properties that is predicted over the simulation snapshots. We use interpolation to determine the positions of galaxies at epochs intermediate to the simulation snapshots, which represents an improvement over previous work. We have shown that our adopted interpolation

scheme leads to accurate predictions for real space galaxy clustering down to scales well within the one-halo regime.

We can summarise our method for constructing lightcone catalogues as follows:

1. Populate the dark matter halos in the snapshot outputs of a cosmological N-body simulation with galaxies using a physical model of galaxy formation, giving populations of galaxies at a range of cosmic epochs. Here we use the dark matter halos from the *Millennium Simulation* (Springel et al., 2005), which we populate with galaxies, whose positions and properties are calculated using the `GALFORM` semi-analytical model.

2. Position an observer within the simulation box. Replicate the simulation box to span a cosmological volume that is of sufficient size to encompass the galaxy survey that we wish to mimic.

3. For replication of the box, use adjacent pairs of simulation snapshots to determine the epoch at which each galaxy enters the observer's past lightcone. Use interpolation to determine the corresponding position of the galaxy at this epoch. Reject all galaxies that enter the observer's lightcone at a position outside of the solid-angle of the galaxy survey.

4. Assign each galaxy that enters the lightcone the intrinsic properties that the galaxy had at the lowest redshift snapshot prior to the galaxy entering the lightcone. Use the position of the galaxy to convert luminosities and absolute magnitudes into observed fluxes and apparent magnitudes. Reject all galaxies that fall outside of the flux limits which define the galaxy survey.

Our approach has a number of attractive features. First, we use a physical model of galaxy formation which makes *ab initio* predictions. This means that we can build mocks for epochs or selections which are currently unprobed. Empirical approaches are not able to do this, as they depend on the existence of observations. Second, our construction method is generic and is not tied to a particular choice of N-body simulation or semi-analytic model. As better N-body simulations or more accurate galaxy formation models become available, our method can still be used. Third, the semi-analytic model that we have used has a unique multi-wavelength capability, which means that we can mimic a wide variety of galaxy surveys spanning a large range in wavelength.

Chapter 4

The BzK colour selection

4.1 Introduction: the redshift desert

Our current understanding of galaxy formation suggests that the epoch spanning the redshift interval $1 \lesssim z \lesssim 3$ marks an important period during which many of the most massive galaxies that we see today formed the vast majority of their stellar mass. Madau et al. (1998) were the first to demonstrate that this period corresponds to the peak in the evolution of the star-formation rate density of the Universe. This has been confirmed by many subsequent follow-up observations (e.g. Dickinson et al., 2003).

Unfortunately this epoch lies within the 'redshift desert', a broad redshift interval where the spectroscopic measurement of galaxy redshifts is difficult due to the lack of strong spectral features at optical wavelengths. Only recently, with the development of near-infrared (NIR) spectroscopy, have large galaxy surveys begun to probe this region and to assess the build-up of galaxies over this crucial period (e.g. Franx et al., 2003; van Dokkum et al., 2003). Prior to this, knowledge of the galaxy population in the 'desert' was derived from photometry. This led to the development of colour selection techniques designed to efficiently identify targets within a specific redshift range for spectroscopic follow-up (which is much more expensive). A well-known example of this is the Lyman-break dropout technique, proposed by Steidel et al. (1996, 2003, 2004), which identifies star-forming galaxies at $z \sim 3 - 10$ according to their rest-frame ultraviolet colours and sampling of the Lyman break spectral feature. Other examples include extremely red objects at $z \sim 1$ (Elston et al., 1988; McCarthy, 2004) and distant red galaxies at $z \sim 2$ (Franx et al., 2003).

In this chapter we will use a lightcone mock catalogue, constructed using the technique described in Chapter 3, to consider the effectiveness of one of these colour selections, the BzK technique (Daddi et al., 2004a).

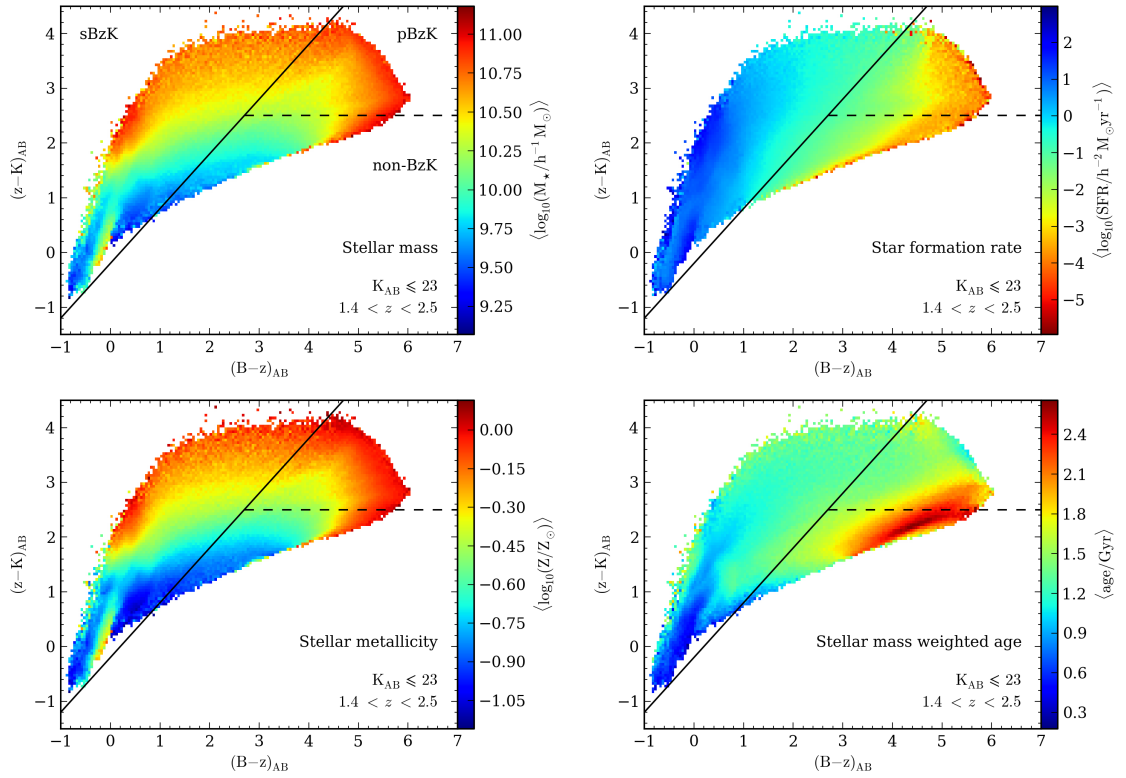

Figure 4.1: The predicted distribution of $K_{AB} \leqslant 23$ galaxies with $1.4 < z < 2.5$ in the BzK colour plane, colour coded, as indicated by the key on the right of each panel, according to the median value in a 2-dimensional colour-colour bin for different galaxy properties: stellar mass (upper left), star-formation rate (upper right), stellar metallicity (lower left) and stellar mass weighted age (lower right). The solid and dashed lines correspond to the BzK criteria used by Daddi et al. (2004a) (see 4.2 for further details). The sBzK and pBzK regions have been labelled in the upper left panel.

4.2 The BzK colour selection

A popular photometric technique, designed to simultaneously identify populations of star-forming and passively evolving galaxies, is the BzK colour-criterion (Daddi et al., 2004a). This approach, which selects galaxies based on their $(B-z)$ and $(z-K)$ colours, is designed to deliver galaxy samples within the redshift range $1.4 < z < 2.5$ that are not biased by the presence of dust or by the age of their stellar populations (Kong et al., 2006; Hayashi et al., 2007, 2009; Grazian et al., 2007; Hartley et al., 2008; Lin et al., 2012).

In Figure 4.1 we show the predicted trends in four galaxy properties across the $(B-z)$ vs. $(z-K)$ colour plane for a sample of $K_{AB} \leqslant 23$ galaxies with $1.4 < z < 2.5$. Galaxies have been placed into 2-dimensional colour-colour bins, which have been colour coded according to the median value of the galaxy property for galaxies in that bin.

Star-forming galaxies at $z > 1.4$ are identified using the criterion

$$\mathrm{BzK} \geqslant -0.2, \tag{4.1}$$

where $\mathrm{BzK} \equiv (\mathrm{z}-\mathrm{K})_{\mathrm{AB}} - (\mathrm{B}-\mathrm{z})_{\mathrm{AB}}$. This condition is indicated by the solid black line in Figure 4.1. Any galaxies lying above this line are considered to be star-forming and are referred to as star-forming BzK, or sBzK, galaxies. The sBzK region has been labelled in the upper left panel of Figure 4.1.

Galaxies lying at $z > 1.4$ that are considered to be passively evolving are identified according to:

$$\mathrm{BzK} < -0.2 \text{ and } (\mathrm{z}-\mathrm{K})_{\mathrm{AB}} > 2.5. \tag{4.2}$$

Galaxies obeying the above criteria are referred to as passive BzK, or pBzK galaxies and populate the region between the solid and dashed lines in Figure 4.1, i.e. the upper right region of the $(\mathrm{B}-\mathrm{z})$ vs. $(\mathrm{z}-\mathrm{K})$ colour plane.

The BzK selection works by using colours that sample key features in the spectral energy distributions of galaxies at $1.4 < z < 2.5$, mainly the rest-frame 4000Å break and the UV continuum slope. Figure 4.2 shows the synthetic spectra for two galaxies at $z = 2$, obtained using the `PEGASE.2` stellar population synthesis code (Fioc & Rocca-Volmerange, 1999). The red lines correspond to a galaxy that is dominated by an old stellar population and thus exhibits a prominent break around 4000Å (in the rest frame), which is created by the build-up of the absorption lines of ionised metals. Between $1.4 < z < 2.5$ the break moves over the observed wavelength range $\sim 9000 - 15000$Å, between the z- and K-bands. In this redshift range we find that the $(\mathrm{z}-\mathrm{K})$ colours of model galaxies become monotonically redder with increasing strength of the 4000Å break.

At $z \leqslant 1.4$, the continuum longwards of the 4000Å break is shifted into the z-band, resulting in galaxies at these redshifts having bluer $(\mathrm{z}-\mathrm{K})$ colours (and, for a fixed B-band flux, redder $(\mathrm{B}-\mathrm{z})$ colours). The BzK criteria are therefore designed to exclude these galaxies, which lie to the right of the solid line and below the dashed line in Figure 4.1. However, as we will see in Section 4.6, the finite width of both the break and the z-band filter, mean that we would expect some contamination to occur, as well as the loss of some galaxies within the target redshift interval, $1.4 < z < 2.5$.

As can be seen in Figure 4.2, the 4000Å break is stronger for galaxies with old stellar populations or high stellar metallicity (Kauffmann et al., 2003a; Kriek et al., 2006, 2011). As such, we would expect old, metal rich galaxies to display redder $(\mathrm{z}-\mathrm{K})$ colours. The lower two panels in Figure 4.1 show the predicted variation of metallicity (left) and stellar

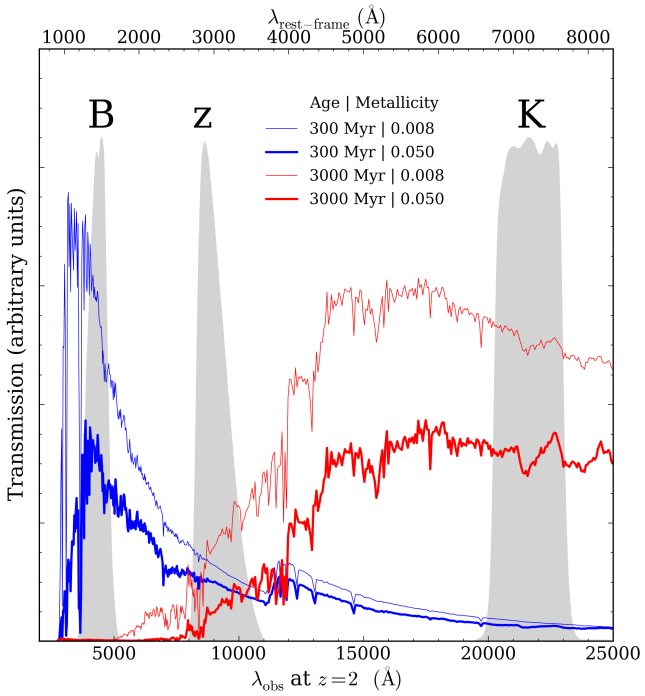

Figure 4.2: The shading shows the transmission curves of the B, z and K filters used by Daddi et al. (2004a). Also shown are the synthetic spectra (plotted as luminosity per unit wavelength) for two galaxies at $z = 2$. The spectra were obtained using `PEGASE.2` (Fioc & Rocca-Volmerange, 1999), assuming a Kennicutt (1983) IMF, and a single instantaneous burst of star formation. The spectra are shown for two different epochs, when the stellar population has an age of 300 Myr (blue) and an age of 3000 Myr (red) and for a sub-solar (thin line) and a super-solar (thick line) metallicity. The flux and transmission units are arbitrary and the spectra have been normalised so as to be visible on similar scales to the transmission curves.

mass weighted age (right) within the (B − z) vs. (z − K) plane. The predicted trends with (z − K) colour agree with the expectations for the variation of the 4000Å break with both age and metallicity. However, these trends are weakened by the effect of dust in young galaxies, which reddens the (z − K) colour.

To isolate young, metal poor galaxies at $1.4 < z < 2.5$ that have not yet developed a strong 4000Å break, another spectral feature is required. Young, star-forming galaxies have steep UV continua due to the presence of bright, young, hot stars. At $z \sim 2$ the UV continuum is shifted into the optical, as shown in Figure 4.2. The presence of the steep UV slope boosts the B-band flux of these galaxies, leading them to have very blue (B − z) colours, as can be seen in the lower right panel of Figure 4.1. The correlation between UV luminosity, due to young stars, and star-formation rate, SFR, suggests that we would also expect a correlation between the SFR of a galaxy and its (B − z) colour for a given K-band limit. Such a trend is clearly visible in the upper right panel of Figure 4.1.

4.3 A BzK mock catalogue

The key question we wish to address is: are the properties of the bright galaxies identified by this selection technique representative of the overall population with $1.4 < z < 2.5$, or are we really just seeing a special subclass of galaxies? If the latter is true, is this simply because current observations have not been sufficiently deep to see the fainter, more representative galaxies or is the BzK criterion somehow biased towards selecting a subset of the galaxy population? A galaxy mock catalogue is a vital resource in helping to answer these questions by allowing an assessment of the effects of observational selection on the completeness of the galaxy sample as well as an examination of whether or not the BzK criterion is sensitive to the intrinsic properties of galaxies.

We therefore construct a lightcone mock catalogue by selecting all galaxies with $K_{AB} \leqslant 24$ over an area of $100\,\mathrm{deg}^2$. The `GALFORM` model used is that of Bower et al. (2006).

Since the BzK technique is used to select a subsample of K-band (or B-band) selected galaxies, we first inspect the predicted total number counts of all galaxies in the B and K-bands in the mock catalogue, which are shown by the solid lines in Figure 4.3. In both bands, the `GALFORM` mock catalogue provides a reasonable match to the observed counts. The B-band counts are in excellent agreement with the observed numbers, though turn over at $B_{AB} \sim 25$ due to the $K_{AB} \leqslant 24$ selection used to construct the lightcone.

As a sanity check, the dotted lines in Figure 4.3 show the differential number counts obtained by integrating the `GALFORM` galaxy luminosity function over co-moving volume. The excellent agreement between the counts computed using the luminosity functions directly from the snapshot outputs and those from the lightcone demonstrates the success of the magnitude interpolation scheme used to create the lightcone. In the right-hand panel of Figure 4.3, the counts computed from the luminosity function diverge from the predicted counts in the lightcone due to the K-band limit used to construct the lightcone (whereas the integral over the luminosity function is independent of this limit).

4.4 Predicted numbers of BzK galaxies

Before considering the effectiveness of the BzK technique, we first assess whether the `GALFORM` mock catalogue is able to predict the correct number of BzK galaxies.

In the left-hand panel of Figure 4.4 we show the number counts of all BzK galaxies with $K_{AB} \leqslant 24$ from the mock catalogue (solid line). Overall the mock provides a reasonable match to the observed counts. At faint magnitudes ($K_{AB} \gtrsim 22$), the turnover in the

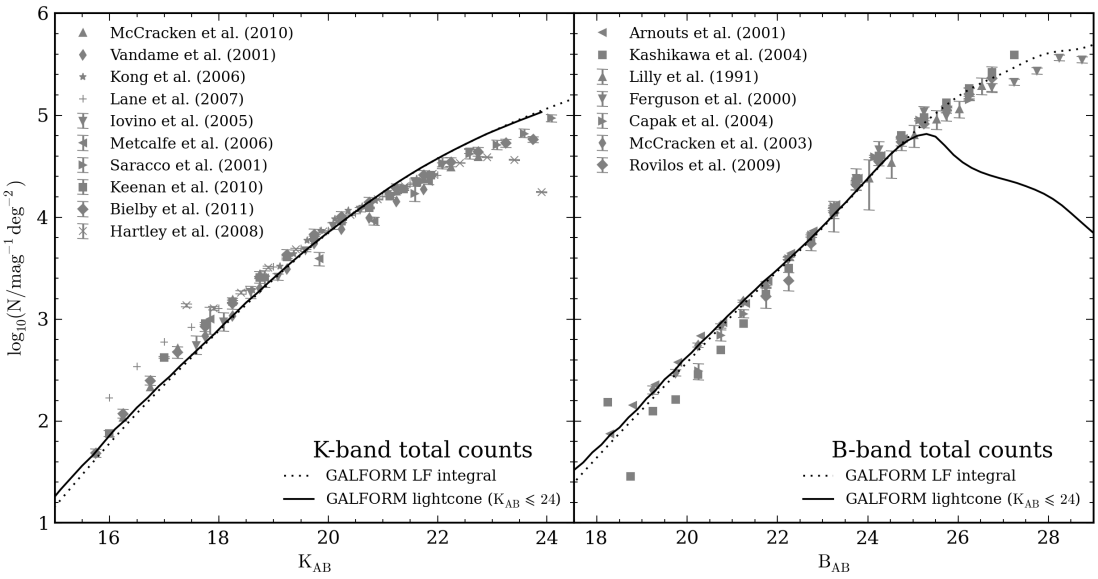

Figure 4.3: Predicted K_{AB}-band (left) and B_{AB}-band (right) differential number counts for all $K_{AB} \leqslant 24$ selected galaxies in the lightcone constructed using the Bower et al. (2006) model (solid lines). The dotted lines show the number counts calculated by integrating the `GALFORM` galaxy luminosity function over comoving volume. The latter uses a single band limit, hence the discrepancy with the counts from the lightcone in the B-band. Also shown are observationally estimated K-band number counts from Saracco et al. (2001); Vandame et al. (2001); Iovino et al. (2005); Metcalfe et al. (2006); Kong et al. (2006); Lane et al. (2007); Hartley et al. (2008); Keenan et al. (2010); McCracken et al. (2010); Bielby et al. (2012) and B-band number counts from Lilly et al. (1991); Ferguson et al. (2000); Arnouts et al. (2001); McCracken et al. (2003); Kashikawa et al. (2004); Capak et al. (2004); Rovilos et al. (2009).

observed BzK counts is sharper than predicted. However, in this region the observations could be incomplete. The closest agreement between the predictions and observations occurs for $K_{AB} \sim 20.5 - 22.0$, where there is a clear change of slope in both the observations and the `GALFORM` predictions. At $K_{AB} \sim 21$, approximately 1/6 of both observed and predicted K-band selected galaxies are also BzK galaxies. Brightwards of $K_{AB} \sim 19.5$, the predicted BzK counts exceed the counts for K-band selected galaxies within $1.4 < z < 2.5$ (shown by the dotted line) due to low redshift interlopers (see Section 4.6.2).

We now consider the predicted number counts for the subsamples of sBzK and pBzK galaxies, shown in the middle and right panels of Figure 4.4. For faint fluxes ($K_{AB} \gtrsim 21$), sBzK galaxies, both observed and predicted, dominate the BzK population due to the turnover in the pBzK counts that can be clearly seen in the right-hand panel of Figure 4.4.

The `GALFORM` sBzK number counts show a good overall agreement with the observations. However, the model over-predicts the number of the faintest sBzK galaxies. This

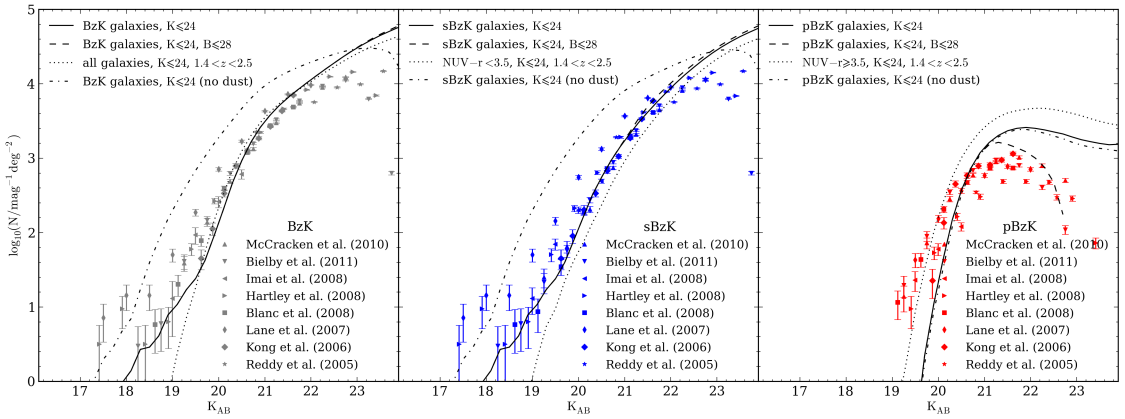

Figure 4.4: Predicted $K_{\rm AB}$-band differential number counts for all BzK (left), sBzK (middle) and pBzK selected galaxies (right) with $K_{\rm AB} \leqslant 24$ in the lightcone catalogue (solid lines). The dashed lines show the predicted number counts when a B-band detection limit of $\rm B_{AB} \leqslant 28$ is considered in addition to the K-band limit (see Section 4.4, in the left-hand panel, this line is underneath the solid one). The dot-dashed lines show the BzK number counts when extinction due to dust is omitted (see Section4.7.5). In the left-hand panel, the dotted line shows the counts for all $K_{\rm AB} \leqslant 24$ selected galaxies within $1.4 < z < 2.5$. In the middle panel the dotted line corresponds to the counts of galaxies with $\rm NUV - r < 3.5$, $K_{\rm AB} \leqslant 24$ and $1.4 < z < 2.5$ and in the right-hand panel the dotted line corresponds to the counts of galaxies with $\rm NUV - r \geqslant 3.5$, $K_{\rm AB} \leqslant 24$ and $1.4 < z < 2.5$ (see Section 4.4 for further details on the colour cut). Also shown are observationally estimated number counts from Reddy et al. (2005); Kong et al. (2006); Lane et al. (2007); Blanc et al. (2008); Hartley et al. (2008); Imai et al. (2008); McCracken et al. (2010); Bielby et al. (2012).

may simply be the result of the observed sBzK counts becoming incomplete at faint magnitudes. The predicted number counts of pBzK galaxies are in reasonable agreement with observations in the range $19.8 \lesssim \rm K_{AB} \lesssim 20.8$. However, the model under-predicts the number of brighter pBzK galaxies and over-predicts the number of fainter galaxies. This mismatch between semi-analytical predictions and observed pBzK number counts has previously been shown by McCracken et al. (2010), who compared their observational counts to the predictions of the mock catalogues of Kitzbichler & White (2007). Moreover, the Kitzbichler & White model gave a poorer match to the observed sBzK counts than we find.

The turnover at faint magnitudes in the counts of pBzK selected galaxies has been reported by several authors (e.g. Lane et al., 2007; Hartley et al., 2008; McCracken et al., 2010; Bielby et al., 2012). The model also displays a turnover in the pBzK counts, but at $\rm K_{AB} \sim 21$, approximately 1 magnitude fainter than in the data. Both Hartley et al. (2008) and McCracken et al. (2010) propose that limited B-band photometry is responsible for

the turnover. Hartley et al. showed that reclassifying ~ 34 per cent of their sBzK galaxies as pBzK galaxies, would be sufficient to remove the turnover (we will return to this point in Section 4.6.3). We demonstrate the impact of the depth of the B-band photometry by recalculating the predicted counts of BzK, sBzK and pBzK galaxies assuming a B-band detection limit of $B_{AB} = 28$ (in addition to the K-band limit of $K_{AB} \leqslant 24$). Any galaxy in the mock catalogue with a B-band magnitude fainter than this is assumed to be undetected in B and its $(B - z)$ colour is calculated assuming $B_{AB} = 28$. This is the approach typically used in observational catalogues to estimate the colours of objects that are undetected in a band. The effect this has on the counts is shown by the dashed lines in Figure 4.4. Although the predicted pBzK counts are still not in full agreement with the data, applying a B-band limit has reduced the mismatch. With the B-band limit applied, ~ 50 per cent of the pBzK galaxies are relabelled as sBzK galaxies preferentially at the faintest K-band magnitudes. This supports the conclusion that shallow B-band photometry contributes to the turnover. Hartley et al. and McCracken et al. both observed the turnover for K-band limited samples down to $K_{AB} \leqslant 23.5$ and $K_{AB} \leqslant 23$ respectively, with B-band detection limits of $B_{AB,\,lim} = 28.4$ and $B_{AB,\,lim} = 29.1$ respectively. The predicted excess of faint pBzK galaxies is also partially a result of the Bower et al. (2006) model predicting too many red galaxies. A substantial fraction of pBzK galaxies are satellites. These galaxies could be too red due to the treatment of gas stripping in satellite subhalos (see Font et al., 2008). If we plot the predicted pBzK number counts considering only central galaxies (without applying any B-band detection limit), we find that the predicted excess of faint pBzK galaxies is reduced, leading to excellent agreement with the observed counts.

The BzK criteria is not the only technique used observationally to classify galaxies as star-forming at $z \sim 2$. The near-UV/optical colour, $(NUV - r)$, can also be used to separate star-forming and passive galaxies. Following Ilbert et al. (2010), we divide the K-band selected `GALFORM` galaxies lying within $1.4 < z < 2.5$ into star-forming galaxies (i.e. those with $NUV - r < 3.5$) and passively evolving galaxies ($NUV - r \geqslant 3.5$), and calculate their number counts. As shown in the middle panel of Figure 4.4, the predicted sBzK counts are consistently somewhat higher than those predicted for galaxies with a blue $(NUV - r)$ colour in $1.4 < z < 2.5$. We note, however, that low-redshift contamination will exaggerate the counts in the brightest bins. The predicted number counts of pBzK galaxies are systematically below the counts of passive galaxies estimated using the $(NUV - r)$ colour. This highlights the sensitivity of the separation of galaxies into star-forming and passively evolving classes to the precise colour criteria used.

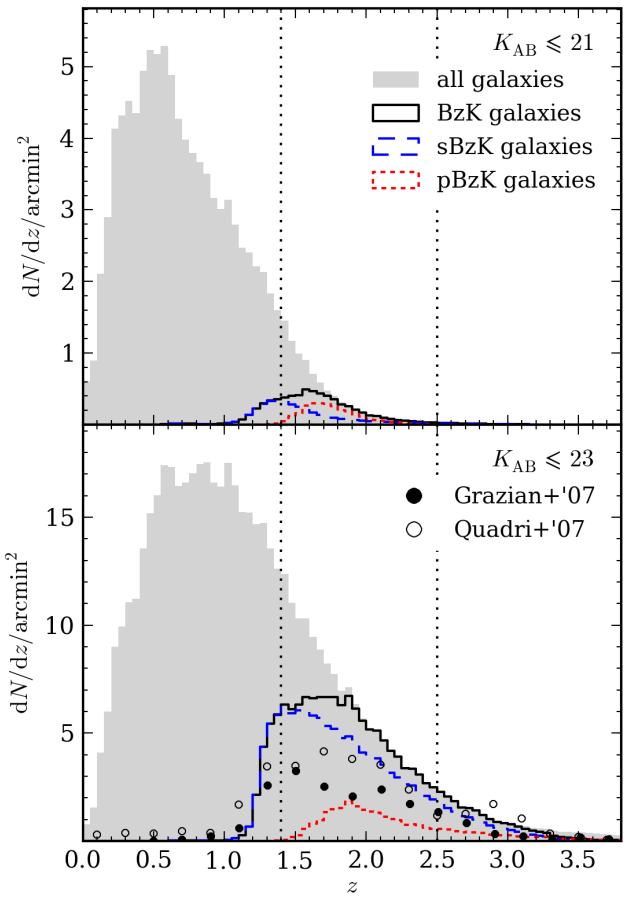

Figure 4.5: The predicted redshift distributions of BzK selected galaxies in the mock catalogue (black solid line) for two different K-band flux limits: $K_{AB} \leqslant 21$ (top) and $K_{AB} \leqslant 23$ (bottom). For comparison, the redshift distribution of all K-band selected mock galaxies down to these limits is shown by the grey shaded region. The limits of the redshift range which the BzK technique was designed to probe are indicated by the vertical dotted lines. Blue dashed and red dotted histograms show the redshift distributions of sBzK and pBzK galaxies respectively. In the bottom panel are plotted observed redshift distributions for BzK galaxy samples with $K_{AB} \leqslant 23.8$ and $K_{AB} \leqslant 22.9$ from Grazian et al. (2007) and Quadri et al. (2007) respectively.

The mock catalogue is able to reproduce the combined number counts of all BzK galaxies, as well as providing reasonable agreement with the counts of K-band selected galaxies within $1.4 < z < 2.5$. Although the model is able to reproduce the predicted number counts of sBzK galaxies (which dominate the BzK population), it is unable to reproduce the predicted number counts of pBzK galaxies.

4.5 Predicted redshift distribution of BzK galaxies

Since the primary aim of the BzK technique is to select galaxies within the redshift interval $1.4 < z < 2.5$, we begin by examining the predicted redshift distribution of the GALFORM mock catalogue and the distribution of those galaxies identified as BzK galaxies.

In Figure 4.5 we show the predicted redshift distributions for BzK galaxies in the GALFORM mock catalogue for two example K-band flux limits, $K_{AB} \leqslant 21$ and 23. For comparison, we also show the redshift distribution for all model galaxies brighter than the stated K-band limit, and use vertical dotted lines to indicate the redshift range, $1.4 < z < 2.5$, which the BzK technique is designed to probe.

It is clear from Figure 4.5 that BzK galaxies probe the high redshift tail of the redshift distribution of K-selected galaxies. For example, the predicted median redshift of the $K_{AB} \leqslant 23$ sample is $z_{med} \sim 1.2$, while the BzK subsample has a higher median redshift of $z_{med} \sim 1.9$. Moreover ~ 98 per cent of the GALFORM galaxies within $2.0 \lesssim z < 2.5$ are selected as BzK galaxies. However, in the redshift range $1.4 < z \lesssim 2$, the fraction of galaxies selected by the BzK technique decreases with decreasing redshift. For $K_{AB} \leqslant 21$ the fraction of galaxies at $z = 1.4$ that are recovered is $\sim 20 - 25$ per cent, compared to ~ 50 per cent for $K_{AB} \leqslant 23$. For $K_{AB} \leqslant 21$ the fraction of galaxies recovered reaches 50 per cent at $z \sim 1.55$.

In the lower panel of Figure 4.5 we compare the redshift distribution for our BzK selected galaxy sample with observed photometric redshift distributions from Grazian et al. (2007) and Quadri et al. (2007), selected with $K_{AB} \leqslant 23.8$ and $K_{AB} \leqslant 22.9$ respectively. The predicted BzK redshift distribution has a median redshift, $z_{med} \simeq 1.8$, that is consistent with the median redshifts of the observed distributions, $1.7 \lesssim z_{med} \lesssim 1.9$. As we have seen in the left-hand panel of Figure 4.4, the GALFORM model over-predicts the number counts of BzK galaxies at $K_{AB} = 23$ and so, understandably, for all redshift bins within $1.4 < z < 2.5$, the mock catalogue predicts a greater number of BzK galaxies than is observed.

We can see from Figure 4.5 that the redshift distribution of sBzK galaxies consistently peaks at lower redshifts than the pBzK distribution. This can also be seen in Figure 4.6, which shows the predicted large-scale distribution of $K_{AB} \leqslant 21$ predicted galaxies and the subsamples of sBzK and pBzK galaxies. Figure 4.6 shows that, while sBzK galaxies can be selected at redshifts down to $z \sim 0$, pBzK galaxies only start to appear at $z \sim 1.4$. In Figure 4.6 we can also see that at $z \sim 2$ the pBzK galaxies appear to trace filamentary

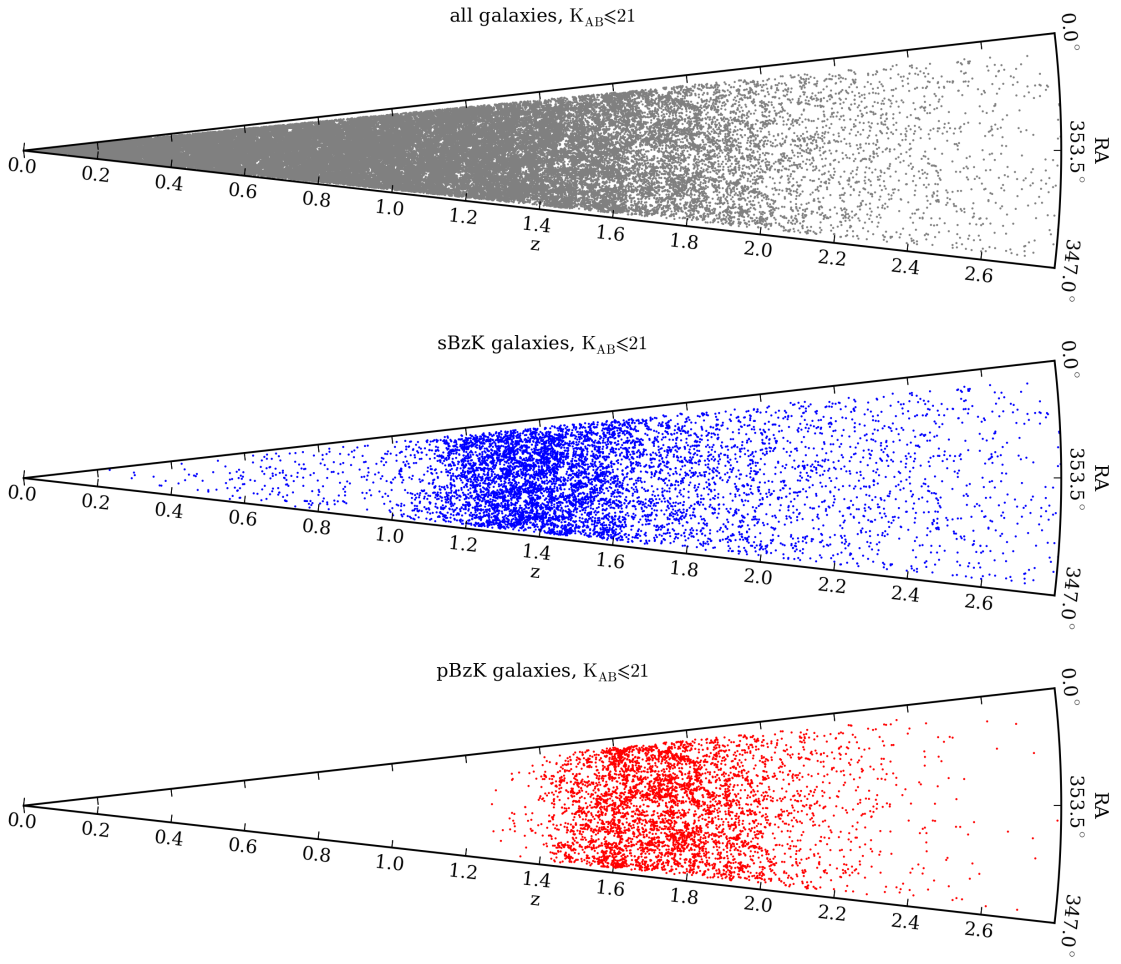

Figure 4.6: Wedge plots showing a slice in redshift and right ascension, 1° wide in declination, of the predicted distribution of all galaxies with $K_{AB} \leqslant 21$ (top) and the subsamples of sBzKs (middle) and pBzKs (bottom).

structures compared to the sBzK galaxies, which appear to be less clustered. Only for fainter limits of $K_{AB} \lesssim 23$, do sBzK galaxies begin to trace the filamentary structure at $z \sim 2$. This suggests that the predicted spatial clustering of pBzK galaxies is stronger than that for sBzKs, in agreement with observations (e.g. Kong et al., 2006; Hartley et al., 2008).

4.6 Efficiency of the BzK selection

The BzK technique was designed to select galaxies within $1.4 < z < 2.5$ and to separate them into star-forming and passively evolving subsamples. To assess the effectiveness with which the BzK technique achieves these goals, we consider the completeness (Section 4.6.1) and contamination (Section 4.6.2) of a BzK galaxy sample selected from the `GALFORM` mock

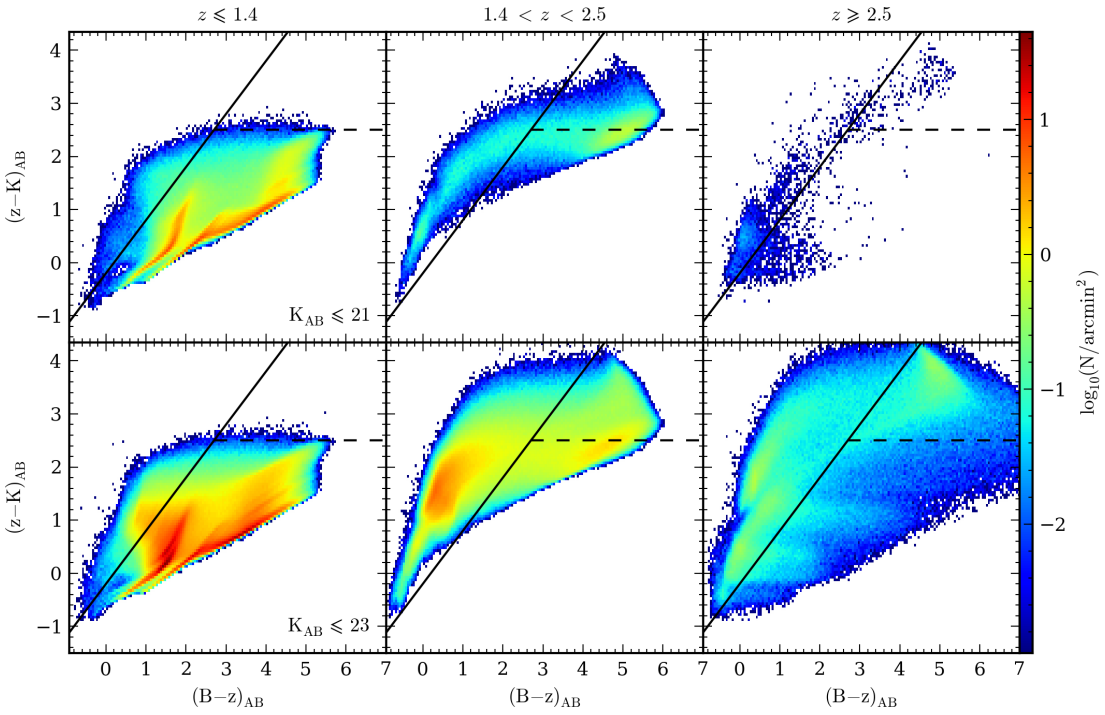

Figure 4.7: The distribution of synthetic galaxies in the BzK colour plane for two K-band flux limits, $K_{AB} \leqslant 21$ (top row) and $K_{AB} \leqslant 23$ (bottom row). The columns correspond to three different redshift ranges: $z \leqslant 1.4$ (left), $1.4 < z < 2.5$ (the redshift interval which the BzK technique was designed to select, middle) and $z \geqslant 2.5$ (right). The black solid line and dashed line correspond to the sBzK and pBzK cuts of Daddi et al. (2004a) respectively. The colour shading indicates the surface density of galaxies on the mock sky, as shown by the scale on the right-hand side.

catalogue.

4.6.1 K-band completeness

In this section we explore the fraction of K-band selected galaxies at $1.4 < z < 2.5$ that are actually picked up when using the BzK selection technique for galaxies in the GALFORM mock catalogue. For this purpose we compare the predicted number counts of BzK galaxies, presented in the left-hand panel of Figure 4.4, with the total number counts of K_{AB}-band selected galaxies that lie within the target redshift range, shown by the dotted line in the same panel of Figure 4.4. Faintwards of $K_{AB} \sim 19.5$ the predicted BzK counts are in good agreement with the counts of $1.4 < z < 2.5$ galaxies, indicating that the BzK selection is an effective probe of the galaxy population at this epoch.

In Figure 4.7 we show the $(B-z)$ vs. $(z-K)$ plane populated by GALFORM galaxies within three different redshift regimes, $z \leqslant 1.4$ (left column), $1.4 < z < 2.5$ (middle

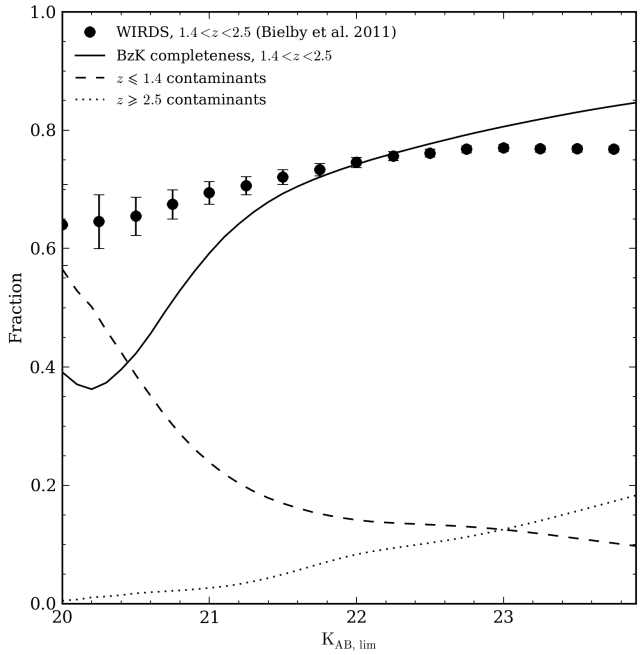

Figure 4.8: The efficiency of the BzK selection as a function of K-band limiting magnitude, $\mathrm{K_{AB,lim}}$. The solid line shows the predicted fraction of `GALFORM` galaxies within $1.4 < z < 2.5$, with $\mathrm{K_{AB}} \leqslant \mathrm{K_{AB,lim}}$, that are identified as BzK galaxies. The filled circles correspond to completeness estimates for observed galaxies in the WIRCam Deep Survey (WIRDS, Bielby et al., 2012) that have been calculated in the same way as the `GALFORM` predictions. The error bars shown correspond to Poisson errors. The dashed and dotted lines show the predicted fraction of interlopers at $z \leqslant 1.4$ and $z \geqslant 2.5$ respectively, as a function of K-band limiting magnitude.

column) and $z \geqslant 2.5$ (right column). The distribution is shown for our two example K-band flux limits: $\mathrm{K_{AB}} \leqslant 21$ (top row) and $\mathrm{K_{AB}} \leqslant 23$ (bottom row). We define the completeness of the BzK technique as the fraction of all galaxies in $1.4 < z < 2.5$ that lie in either of the BzK regions in the $(\mathrm{B}-\mathrm{z})$ vs. $(\mathrm{z}-\mathrm{K})$ plane. About a quarter of the galaxies brighter than $\mathrm{K_{AB}} = 21$ within $1.4 < z < 2.5$ lie outside of the BzK regions. The distribution has two clear peaks, one at $(\mathrm{B}-\mathrm{z}) \sim 0$, which we will refer to as the star-forming peak, and the other at $(\mathrm{B}-\mathrm{z}) \sim 5$, which we will refer to as the passively evolving peak. The star-forming peak falls well within the sBzK region, while the passively evolving peak lies just outside the pBzK region. This would explain the under-prediction of the pBzK number counts for $\mathrm{K_{AB}} \lesssim 20$. However, for $\mathrm{K_{AB}} \leqslant 23$, the star-forming peak dominates the galaxy population suggesting that for fainter K-band limits the BzK selection provides a more complete sample of the $1.4 < z < 2.5$ galaxy population.

The completeness of the BzK technique, as a function of the limiting K-band magnitude of the galaxy sample, is shown in Figure 4.8. Here, the data points show the

BzK completeness estimates from Bielby et al. (2012), who applied the BzK selection to an input catalogue of $\sim$ 1.8 million K-band galaxies, with photometric redshifts ($\sigma_{\Delta z/(1+z)} \lesssim 0.03$), from the WIRCam Deep Survey (WIRDS). We can clearly see in Figure 4.8 that the BzK completeness increases with fainter K-band limiting magnitude. The same trend is seen for the completeness predictions for `GALFORM` galaxies, shown by the solid line, with $\sim$ 55, $\sim$ 73 and $\sim$ 80 per cent of $1.4 < z < 2.5$ galaxies being recovered for $K_{AB} \leqslant 21$, 22 and 23 respectively. Therefore, for faint K-band limits ($K_{AB,lim} \gtrsim 22$), the BzK technique is consistently selecting 75 to 80 per cent of the galaxies within $1.4 < z < 2.5$. However, for a very bright limit of $K_{AB} \leqslant 20$ the technique identifies less than half of the galaxy population within the target redshift range. For $21 \lesssim K_{AB,lim} \lesssim 22$ the completeness estimates from the `GALFORM` mock catalogue are in very good agreement with the WIRDS estimates.

4.6.2 Contamination

We now explore the predicted numbers of galaxies outside the redshift range, $1.4 < z < 2.5$, that are picked up by the BzK selection. As we can see from Figure 4.7, it is not possible to ever have a sample of BzK selected galaxies that is entirely free of contamination from interlopers with low, $z \leqslant 1.4$, or high redshift, $z \geqslant 2.5$, which have been classified as BzK galaxies. The left-hand column of Figure 4.7 shows that low redshift interlopers are typically classified as sBzK galaxies, while the $(z - K)$ cut used in Eq. (4.2) successfully eliminates low redshift pBzK galaxies. High redshift interlopers, shown in the right-hand column of Figure 4.7, appear to be more evenly distributed between the sBzK and pBzK regions and thus more difficult to remove.

The fractions of low and high redshift interlopers, as a function of K-band limiting magnitude, are shown in Figure 4.8 by the dashed and dotted lines respectively. From Figure 4.8 we can see that by applying a bright $K_{AB} \leqslant 20$ selection to our `GALFORM` mock catalogue, the BzK technique selects approximately equal numbers of galaxies with $1.4 < z < 2.5$ and $z \leqslant 1.4$. Pushing the K-band selection limit to fainter magnitudes leads to a decrease in the fraction of low redshift contamination as an increasing number of galaxies within $1.4 < z < 2.5$ become visible at fainter K-band limits. Figure 4.5 shows clearly how the redshift distribution of BzK galaxies develops a sharper low redshift cut-off as the flux limit is made fainter. By $K_{AB,lim} \sim 21.5$, the low redshift contamination has fallen to $\sim$ 18 per cent. For fainter flux limits the low redshift contamination decreases much more slowly, reaching $\sim$ 10 per cent by $K_{AB,lim} \sim 24$.

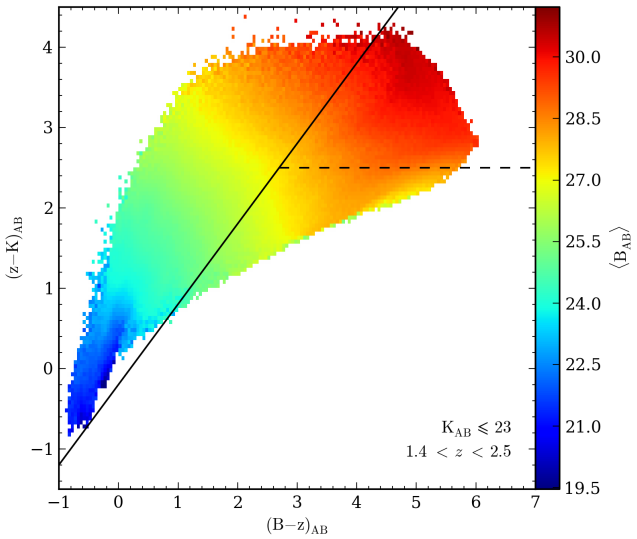

Figure 4.9: The variation in the median B-band apparent magnitude of galaxies with position in the (B − z) vs. (z − K) plane. The distribution shown corresponds to `GALFORM` galaxies, within $1.4 < z < 2.5$, selected to have $K_{AB} \leqslant 23$ and placed into 2-dimensional bins spanning the BzK colour-colour space. The bins are coloured according to the median B-band magnitude of the galaxies in that bin, as shown by the colourbar.

As expected, the fraction of high redshift interlopers increases steadily with increasingly faint limiting K-band magnitude, though it stays well below $\sim$ 20 per cent. By $K_{AB,lim} \sim 23.2$, the contribution from low redshift and high redshift contamination is approximately equal at $\sim$ 12 per cent, with high redshift interlopers dominating the contamination at fainter limiting magnitudes.

4.6.3 Dependence on B-band depth

As we have seen in Section 4.4, there is evidence that the ability of the BzK technique to distinguish between star-forming and passive galaxies within $1.4 < z < 2.5$ is dependent upon the B-band depth of the galaxy sample. For example, Grazian et al. (2007) determined that 22 per cent of their sample of sBzK galaxies had SEDs typical of passive galaxies rather than star-forming galaxies. A significant number of these galaxies were undetected in the B-band and had their (B − z) colours estimated using a 1σ B-band upper limit, which resulted in their (B − z) colours being too blue. Grazian et al. concluded that, for faint K-band selected galaxies with very red (z − K) colours, a lack of deep B-band photometry will lead to many pBzK galaxies being incorrectly classified as sBzK galaxies.

In Figure 4.9, we show the variation of the median B-band apparent magnitude with position in the $(B-z)$ vs. $(z-K)$ plane for $K_{AB} \leqslant 23$ galaxies in the GALFORM mock catalogue. The trend towards fainter B-band magnitudes for redder $(B-z)$ and $(z-K)$ colours is immediately clear and supports the need for deep B-band photometry to probe the faint pBzK population.

We apply a B-band detection limit of $B_{AB,lim} \leqslant 26$ to a $K_{AB} \leqslant 23$ sample of galaxies, by assuming that galaxies with B-band magnitudes fainter than $B_{AB,lim}$ are undetected and so have $B_{AB} = B_{AB,lim}$. By doing this, we find that only ~ 0.3 per cent of galaxies within $1.4 < z < 2.5$ are classified as pBzK galaxies. Making the B-band limit fainter leads to a larger fraction of pBzK galaxies. With upper limits of $B_{AB,lim} \leqslant 27$, 28 and 29 we find that ~ 3, ~ 9 and ~ 15 per cent of $1.4 < z < 2.5$ galaxies respectively are classified as pBzK galaxies. An upper limit of $B_{AB,lim} \leqslant 30$ leads to the same number of pBzK galaxies being recovered (~ 16 per cent) as when applying the $K_{AB} \leqslant 23$ selection in isolation.

A bright B-band limit will also lead to galaxies that should not be classified as BzK galaxies being scattered into the sBzK region of the $(B-z)$ vs. $(z-K)$ plane. As we have seen in Section 4.6.1, for a $K_{AB} \leqslant 23$ selected galaxy sample, the BzK technique selects ~ 80 per cent of $1.4 < z < 2.5$ galaxies. If we apply B-band detection limits of $B_{AB,lim} \leqslant 26$, 27 and 28 we find that the BzK technique selects ~ 95, ~ 87 and ~ 80 per cent of galaxies within $1.4 < z < 2.5$ respectively.

We conclude that adopting a fainter B-band limit should improve the ability of the BzK technique to distinguish between star-forming and passively evolving galaxies.

4.7 The predicted properties of BzK galaxies

Early studies of $K_{AB} \lesssim 22$[1] BzK-selected galaxies revealed them to typically have large stellar masses, $\sim 10^{11} h^{-1} M_{\odot}$, and, in the case of those labelled as star-forming, high star-formation rates, $\sim 100 h^{-1} M_{\odot} yr^{-1}$ (Daddi et al., 2004a,b, 2005a,b; Reddy et al., 2005; Kong et al., 2006). Such properties, combined with high metallicities (e.g. Daddi et al., 2004a; Hayashi et al., 2009) and indications that these systems are strongly clustered (Hayashi et al., 2007; Hartley et al., 2008; Blanc et al., 2008), have led many authors to speculate that BzK galaxies are the high-redshift precursors of massive early-type galaxies

[1]Daddi et al. (2004a) originally used a $K_{Vega} \leqslant 20$ selected sample. We have converted this to the AB system using the K-band conversion from Blanton & Roweis (2007), where $m_{AB} - m_{Vega} = 1.85$.

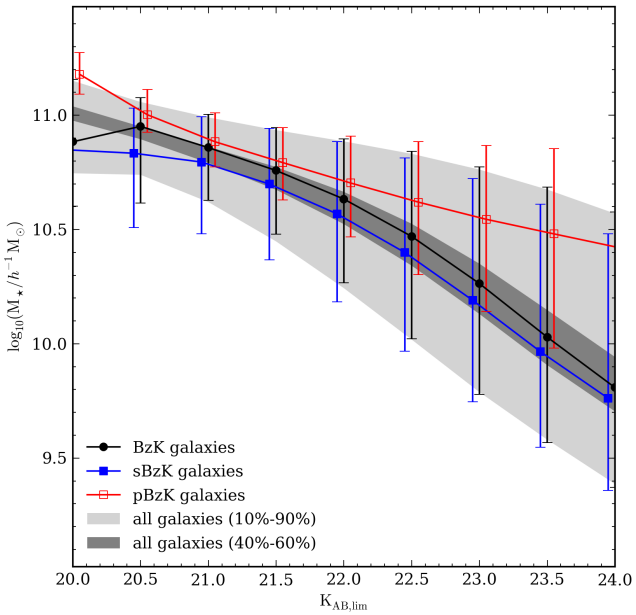

Figure 4.10: The predicted stellar mass of galaxies with redshift $1.4 < z < 2.5$, as a function of K-band limiting magnitude for all BzK galaxies (black circles), sBzK galaxies (blue, filled squares) and pBzK galaxies (red, open squares). Data points correspond to median values and error bars show the 10 and 90 percentiles. For clarity, the data points for the sBzK and pBzK values have been offset horizontally. The light and dark grey regions show the $10-90$ and $40-60$ percentiles for all galaxies brighter than the K-band flux limit (i.e. irrespective of whether they are BzK selected).

found in groups and clusters at the present day. In this section we will examine the typical properties of BzK galaxies as predicted by the `GALFORM` model.

4.7.1 Stellar mass

From the upper left panel of Figure 4.1 we can see that the predicted stellar masses of galaxies in a $K_{AB} \leqslant 23$ selected BzK sample range from $\sim 10^9 h^{-1}M_\odot$ to $\sim 10^{11} h^{-1}M_\odot$, with the more massive galaxies typically having redder $(z-K)$ colours.

We show in Figure 4.10 the distribution of stellar masses for all K-band selected galaxies (within $1.4 < z < 2.5$). The median stellar mass of all BzK selected galaxies is in excellent agreement with the distribution for all K-band selected galaxies for all flux limits fainter than $K_{AB,lim} \sim 21$. Additionally, the 10 and 90 percentiles of the BzK distribution consistently match the 10 and 90 percentiles for the stellar mass distribution of the whole galaxy population.

Early studies of BzK galaxies, using K-band limits of $K_{AB} \lesssim 22$, inferred BzK galaxies

to be very massive, with typical stellar masses[2] of $M_\star \gtrsim 5 \times 10^{10} h^{-1}\mathrm{M}_\odot$ (Daddi et al., 2004b,a, 2005b,a; Reddy et al., 2005; Kong et al., 2006; Blanc et al., 2008). In Figure 4.10 we show the median stellar mass of BzK selected galaxies (within $1.4 < z < 2.5$) as a function of the K-band flux limit. For $\mathrm{K_{AB}} \leqslant 22$ the distribution of BzK stellar masses in the mock catalogue is consistent with observations. Increasing the depth in the K-band leads to a shift in the distribution towards smaller stellar masses, with a median BzK stellar mass of $\sim 10^{10} h^{-1}\mathrm{M}_\odot$ being reached at $\mathrm{K_{AB,lim}} \sim 23.5$.

We also show in Figure 4.10 the breakdown of the distribution into sBzK and pBzK galaxies. It is immediately clear that pBzK galaxies are typically more massive than sBzK galaxies, with the difference between the medians increasing towards fainter K-band limits.

We conclude that BzK selected galaxies appear to provide a representative sample of the galaxy stellar masses at $1.4 < z < 2.5$ and do not appear to be significantly biased towards either very high or low mass galaxies.

4.7.2 Star Formation Rate

As we have already seen in Figure 4.1, there is clear trend in the predicted SFR of galaxies across the $(\mathrm{B} - \mathrm{z})$ vs. $(\mathrm{z} - \mathrm{K})$ plane. In the extremes of the distribution we find that many sBzK selected galaxies are predicted to have SFRs of $\sim 100\, h^{-1}\mathrm{M}_\odot \mathrm{yr}^{-1}$ or more, while many pBzK selected galaxies have SFRs of effectively zero.

In Figure 4.11 we show the distribution of SFRs for BzK, sBzK and pBzK galaxies, as well as for all K-band selected galaxies, as a function of K-band flux limit, in the redshift range $1.4 < z < 2.5$.

For $\mathrm{K_{AB}} \gtrsim 21$, the median SFR for BzK galaxies is in reasonable agreement with the distribution for the whole galaxy population, though is perhaps slightly biased towards higher SFRs. This would, at first, suggest that the BzK selection is missing a fraction of the passive galaxy population, particularly since we have shown in Figure 4.4 that the `GALFORM` mock catalogue matches the number of sBzK galaxies but under-predicts the number of bright pBzK galaxies. It is possible that some fraction of these faint pBzK galaxies are dusty star-forming galaxies that have been mis-classified as being passive. For $\mathrm{K_{AB}} \leqslant 21$, we find that ~ 20 per cent of the pBzK selected galaxies in the `GALFORM` mock

[2]The quoted value for the observed mass has been multiplied by a factor of 1.4 (Fontana et al., 2004) in order to account for the change from Salpeter (1955) IMF, used in observational studies, to the Kennicutt (1983) IMF used for the study presented here.

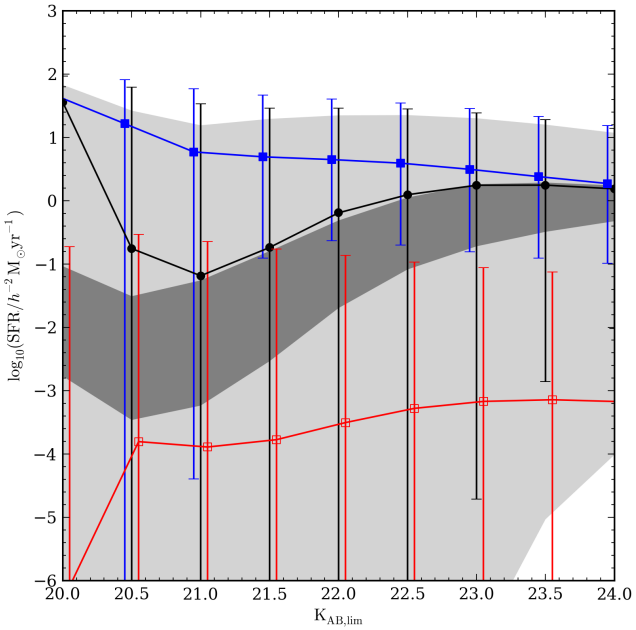

Figure 4.11: The predicted star-formation rate as a function of K-band limiting magnitude for galaxies in $1.4 < z < 2.5$. The symbols and shaded regions are the same as in Figure 4.10.

catalogue have SFRs $\gtrsim 0.1\,h^{-1}\mathrm{M}_\odot\mathrm{yr}^{-1}$. Interestingly, the typical SFR of pBzK galaxies remains approximately constant, at $\sim \left(10^{-4} - 10^{-3}\right)\ h^{-1}\mathrm{M}_\odot\mathrm{yr}^{-1}$, with increasing K-band depth (though the distribution is very broad). For $\mathrm{K}_{\mathrm{AB,lim}} \gtrsim 23$, the typical SFR of sBzK galaxies also appears to remain almost constant at $\sim (1 - 10)\ h^{-1}\mathrm{M}_\odot\mathrm{yr}^{-1}$.

In Figure 4.11, we see that the model predicted median SFR of BzK galaxies with $\mathrm{K}_{\mathrm{AB}} \leqslant 22$ is $\sim 1\,h^{-1}\mathrm{M}_\odot\mathrm{yr}^{-1}$. However, observational studies of BzK galaxies with $\mathrm{K}_{\mathrm{AB}} \lesssim 22$ concluded that many of these bright BzK galaxies were starbursting galaxies, with SFRs of $\sim (50 - 100)\ h^{-2}\mathrm{M}_\odot\mathrm{yr}^{-1}$ (e.g. Daddi et al., 2004a; Kong et al., 2006; Blanc et al., 2008). A possible explanation for this discrepancy is the overly efficient shut down of gas cooling by AGN feedback in the Bower et al. (2006) model, which has been previously suggested by Gonzalez-Perez et al. (2009).

Based upon the model predictions, however, we predict that towards fainter K-band limiting magnitudes the BzK technique is typically selecting galaxies with SFRs that are consistent with the median SFRs of the galaxy population within $1.4 < z < 2.5$.

4.7.3 Metallicity

We have already seen in Section 4.2 that the metallicity of K-band selected galaxies varies with position in the $(\mathrm{B} - \mathrm{z})$ vs. $(\mathrm{z} - \mathrm{K})$ plane. From the lower left-hand panel of

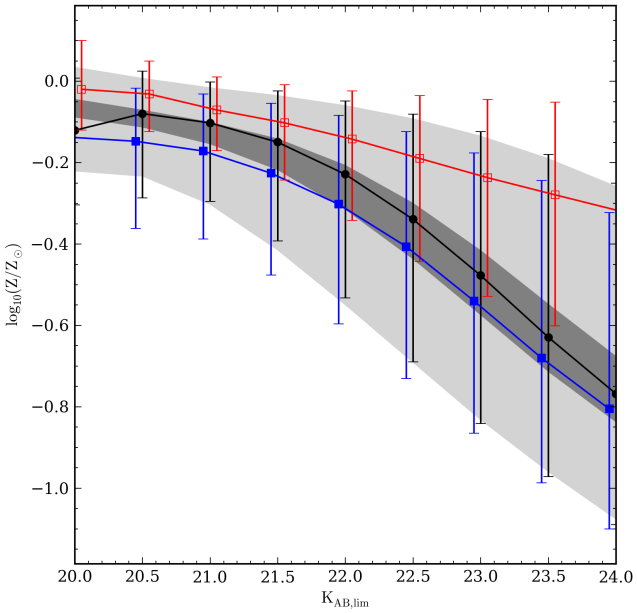

Figure 4.12: The predicted stellar metallicity as a function of K-band limiting magnitude for galaxies in $1.4 < z < 2.5$. The symbols and shaded regions are the same as in Figure 4.10.

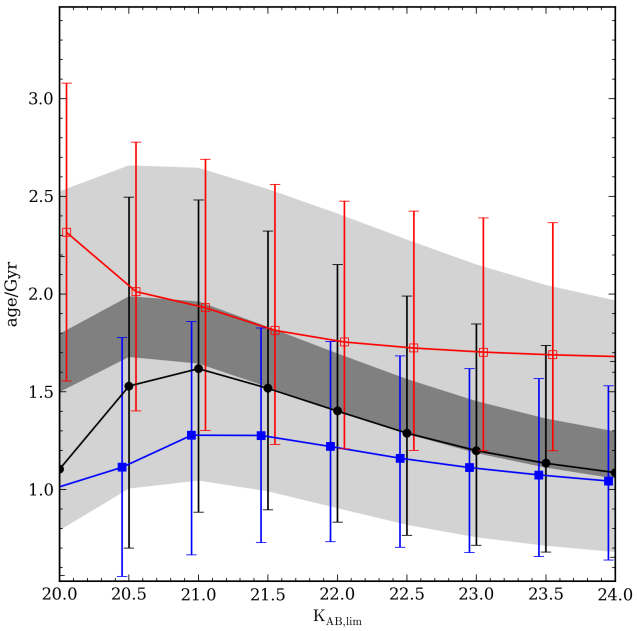

Figure 4.13: The predicted stellar mass weighted age as a function of K-band limiting magnitude for galaxies in $1.4 < z < 2.5$. The symbols and shaded regions are the same as in Figure 4.10.

Figure 4.1 we can see that galaxies with the reddest (z − K) colours (typically pBzK and faint sBzK galaxies) are in general the most metal rich.

In Figure 4.12 we show the metallicity distribution for BzK selected galaxies within $1.4 < z < 2.5$. For all K-band limits considered, the metallicity distribution of BzK selected galaxies is in good agreement with the metallicity distribution for all K-band selected galaxies. The trend in the metallicity distribution as a function of K-band flux limit is very similar to the trend seen in the stellar mass distribution in Figure 4.10. For brighter K-band flux limits, one would predict to recover BzK galaxies with higher metallicities. For the brightest flux limit considered, the median metallicities for BzK galaxies falls below that for all galaxies. As with the stellar mass distribution, this is due to `GALFORM` under-predicting the counts of bright pBzK galaxies, which one would expect to be metal-rich. The distributions for the separate sBzK and pBzK subsets show that for any K-band depth, pBzK galaxies will typically be more metal-rich than sBzK galaxies, though the distributions for the two subsets do overlap.

4.7.4 Age

In the lower right-hand panel of Figure 4.1 we show the median stellar mass weighted age of galaxies in the (B − z) vs. (z − K) plane. We find that the oldest galaxies occupy the region where the density of passive galaxies peaks just below the pBzK region, as can be seen in the middle column of Figure 4.7. The vast majority of the oldest galaxies, with ages above 2 Gyr, that fall outside the pBzK region lie within the redshift interval $1.4 < z < 2.5$. This is due to the finite width of the 4000Å break, which at $z \sim 1.4$ is beginning to enter the response curve of the z-band, thus making the (z − K) colours of these galaxies bluer. Above $z = 2$, all of the galaxies with ages above approximately 1.5 Gyr lie well within the pBzK region on the colour plane. We have checked that the (z − K) colours of the galaxies are not significantly affected by changing between a Kennicutt (1983) and a Salpeter (1955) IMF.

We show in Figure 4.13 the distribution of the stellar mass weighted ages for all K-band selected galaxies and for those that are BzK-selected. Like the distribution of SFRs, the distribution of ages of BzK galaxies is in reasonable agreement with the age distribution for all K-band selected galaxies, though appears to be slightly biased towards younger galaxies. As with the SFR distribution, we see that the typical ages of sBzK and pBzK galaxies remain approximately constant (at $\sim$ 1.1Gyr and $\sim$ 1.7Gyr respectively) for $K_{AB,lim} \gtrsim 22$.

4.7.5 Dust

Reddening due to dust can mimic a large break at 4000Å in the spectra of star-forming $z \lesssim 1.4$ galaxies (Kriek et al., 2006, 2011). However, many authors have argued that the effectiveness of the BzK colour selection is not significantly affected by dust extinction (e.g. Daddi et al., 2004a; Kong et al., 2006; Hayashi et al., 2007; Grazian et al., 2007; Hartley et al., 2008; Hayashi et al., 2009; Lin et al., 2012).

For the K-band limits considered in Figure 4.7 we find that in the presence of dust the distribution of $1.4 < z < 2.5$ galaxies in the $(\mathrm{B}-\mathrm{z})$ vs. $(\mathrm{z}-\mathrm{K})$ plane remains relatively unchanged, aside from an increased scatter in the sBzK galaxy population towards redder $(\mathrm{z}-\mathrm{K})$ colours. We find that for $\mathrm{K_{AB}} \leqslant 23$ the presence of dust reddens the colours of BzK galaxies, within $1.4 < z < 2.5$, by $\Delta(\mathrm{B}-\mathrm{z}) \sim 0.15$ and $\Delta(\mathrm{z}-\mathrm{K}) \sim 0.3$. The presence of dust appears to have a greater effect on the median colours of sBzK galaxies, as we find a negligible change in the median colours of pBzK galaxies. We can see this also in Figure 4.4 where the number counts of sBzK galaxies without dust extinction are boosted by ~ 1 dex, while the pBzK number counts remain the same. The reduction in the sBzK counts when dust extinction is included is likely due to dust reddening the $(\mathrm{B}-\mathrm{z})$ colours of star-forming galaxies (with $(\mathrm{z}-\mathrm{K})_{\mathrm{AB}} < 2.5$) and scattering them out of the sBzK region.

4.7.6 Conclusions & future work

As an illustrative application of our lightcone catalogues we have considered the effectiveness of the BzK colour selection technique, which is designed to isolate galaxies within the redshift range $1.4 < z < 2.5$ (Daddi et al., 2004a). The aim of this exercise was to determine how successful this technique is at isolating galaxies within the target redshift range and whether the galaxies it selects are representative of the target population or a biased subsample.

The `GALFORM` model is able to match reasonably well the K-band number counts of all BzK galaxies, as well as the counts of sBzK galaxies. However, the model under-predicts the number of bright pBzK galaxies and over-predicts the number of faint pBzK galaxies. The latter discrepancy is partially due to the effect of the depth of B-band photometry, but may also be related to the crude estimate of the stripping of gas from satellite galaxies that is carried out in the Bower et al. (2006) model. The BzK technique successfully selects the majority of the galaxy population within $2 \lesssim z < 2.5$ (and possibly out as far

as $z \sim 3$), though is less efficient for $1.4 < z \lesssim 2.0$. Examination of the effectiveness of the BzK technique as a function of K-band limiting magnitude suggests that the technique recovers $\gtrsim 75$ per cent of the $1.4 < z < 2.5$ galaxy population for K-band limits fainter than $K_{AB} \sim 22$. For brighter limits the completeness decreases substantially as the BzK population becomes dominated by low redshift interlopers with $z \leqslant 1.4$. For magnitude limits $K_{AB} \gtrsim 21.5$, the fraction of contamination from BzK galaxies outside $1.4 < z < 2.5$, remains approximately constant at ~ 30 per cent. We have also shown that a variation in the typical B-band magnitude across the BzK plane can lead to the mis-classification of pBzK galaxies as sBzK galaxies if the B-band photometry is of insufficient depth. Finally, we considered the intrinsic properties of BzK galaxies, including their stellar mass, SFR, metallicity and stellar mass weighted age. We find that the properties of BzK galaxies display distributions that are in good agreement with the corresponding distributions for all galaxies with $K_{AB} \gtrsim 20.5$. However, at brighter K-band limits BzK galaxies appear to be less massive, more star-forming, less metal-rich and younger than the overall population. This is likely related to the under-prediction of the bright pBzK number counts. The presence of dust increases the scatter in the colours of (faint) sBzK galaxies, though does not dramatically change the colour distribution of galaxies within $1.4 < z < 2.5$.

We conclude that the BzK colour selection provides a representative sample of the $1.4 < z < 2.5$ galaxy population and is more effective for fainter K-band flux limits. However, the depth of B-band photometry and extinction due to dust may lead to confusion between the sBzK and pBzK subsets.

Possible extensions to this project include:

- Examining the extent to which the effectiveness of the BzK technique distorts the clustering signal of galaxies in the redshift range $1.4 < z < 2.5$. For instance, estimating the angular correlation function, $\omega(\theta)$, for BzK selected galaxies and comparing this to an estimate of $\omega(\theta)$ for all galaxies in $1.4 < z < 2.5$ would reveal the affect of high and low redshift interloper galaxies on the estimates of the $z \sim 2$ clustering signal. Estimating $\omega(\theta)$ for just the BzK galaxies inside the desired redshift range would reveal what effect the incomplete sampling of the $1.4 < z < 2.5$ population has on the clustering strength of these galaxies.

- The analysis that has been carried out in this chapter could be adapted to consider the effectiveness of other colour selection techniques. For example, lightcone cata-

logues could be constructed to investigate the effectiveness of the colour selections for identifying samples of luminous red galaxies (LRGs) or extremely red objects (EROs).

Chapter 5

Using a mock catalogue to calibrate a galaxy group-finding algorithm

5.1 Introduction

Measurement of the evolution and slope of the halo mass function (HMF), and the variation of the halo mass-to-light (M/L) ratio with halo mass, allow one to distinguish between competing cosmological models and examine how the efficiency of galaxy formation varies with halo mass (Eke et al., 1998a, 2004b; Driver et al., 2009). Measurements of the HMF at different epochs allows constraints to be placed on σ_8 (the rms fluctuation in spheres of radius $8\,h^{-1}$Mpc predicted by linear perturbation theory) and $\Omega_{\rm M}$ (Eke et al., 1996, 1998a; Carlberg et al., 1996; Bahcall et al., 1997). Simulations based upon the Λ cold dark matter (ΛCDM) cosmology predict that variation of the M/L ratio as a function of halo mass has a characteristic shape, with a minimum occurring at a mass $\sim 10^{12}h^{-1}{\rm M}_{\odot}$ (a mass similar to that of the Local Group). Given the shape of the halo mass function in the ΛCDM cosmology, it is clear that to transform this into the observed galaxy luminosity function, a strong suppression of the efficiency of galaxy formation in halos of different mass is needed. As a result, the dependence of the M/L ratio on halo mass is a key prediction of galaxy formation models set in the ΛCDM cosmology. Unfortunately, the minimum in the M/L ratio is yet to be cleanly observed (Eke et al., 2004b). Testing the ΛCDM predictions for the HMF and M/L ratio are two of the goals of the next generation of galaxy surveys (such as the Galaxy And Mass Assembly Survey, Driver et al., 2009), which will require accurate measurements of groups and cluster halo masses (e.g. Alpaslan et al., 2012).

Galaxies do not directly trace the matter distribution, but instead are predicted to have a spatial distribution that displays a scale-dependent bias relative to the underlying

mass (Peacock, 1997; Jenkins et al., 1998; Jing et al., 1998; Seljak, 2000; Benson et al., 2000b; Peacock & Smith, 2000). Tools such as the *halo occupation distribution* (HOD, Berlind & Weinberg, 2002) are used to describe the number of galaxies that are expected to be found within halos of a given mass and so allow us to calculate galaxy clustering. Identification of the members of galaxy groups and clusters helps place observational constraints on the HOD (e.g. Capozzi et al., 2012). Furthermore, substantial evidence suggests that various statistics of the galaxy distribution, such as the luminosity function, show a dependence on the mass of the halo hosting a galaxy (e.g. Valotto et al., 1997; De Propris et al., 2003; Eke et al., 2004b; Baldry et al., 2008; Robotham et al., 2010). Observations suggest an environmental dependence of the star formation activity of galaxies in local ($z < 1$) groups and clusters (e.g. Dressler, 1980; Lewis et al., 2002; Goto et al., 2003; Poggianti et al., 2006; Martínez et al., 2006; Cooper et al., 2008; Bouchard et al., 2009; Skibba et al., 2009; Mercurio et al., 2010; van der Wel et al., 2010; Deng et al., 2011). Whether such a dependence exists at higher redshift, however, is still under debate (Elbaz et al., 2007; Ideue et al., 2009; Grützbauch et al., 2011; Sobral et al., 2011; Li et al., 2011; Tanaka et al., 2012; Ideue et al., 2012). Therefore, catalogues of galaxy groups and clusters can provide useful input catalogues for detailed follow-up studies.

Galaxy groups and clusters act as observational probes of the population of dark matter halos and so measurements of the abundance of groups, as a function of redshift, as well as the properties of the galaxies that they host, are vital in helping test the ΛCDM paradigm and improve our understanding of galaxy formation. There has therefore been significant interest in establishing catalogues of galaxy groups and clusters, especially since the development of large volume galaxy redshift surveys (e.g. Merchán & Zandivarez, 2002; Marinoni et al., 2002; Eke et al., 2004a; Yang et al., 2005; Gerke et al., 2005; Crook et al., 2007; Murphy et al., 2011; Robotham et al., 2011; Murphy et al., 2012; Gerke et al., 2012). In the construction of these groups catalogues, several different techniques have been used. One of the most commonly used approaches is the well-tested friends-of-friends (FOF) technique first used by Huchra & Geller (1982). This technique identifies galaxy groups based upon the proximity of the galaxies to one another. It was used by Merchán & Zandivarez (2002) and Eke et al. (2004a) to catalogue groups in the Two-degree Field Galaxy Redshift Survey (2dFGRS), by Berlind et al. (2006) to catalogue groups in the Sloan Digital Sky Survey (SDSS), by Crook et al. (2007) to catalogue groups in the 2MASS Redshift Survey (2MRS) and Robotham et al. (2011) to catalogue groups in the Galaxy And Mass Assembly (GAMA) survey. Murphy et al. (2011) also used the FOF

technique to identify filaments in the 2dFGRS. We will discuss the FOF group finding technique in more detail when we describe our chosen algorithm in Section 5.3.

Yang et al. (2005) and Yang et al. (2007) construct groups catalogues for the 2dFGRS and SDSS using an algorithm that first uses FOF to identify tentative group centres. They then determine the luminosity, line-of-sight velocity dispersion and projected radius of these groups and estimate the mass using a model of the M/L ratio. The final step is to estimate for each galaxy a probability that it is a member of a particular group, based upon the estimated properties of each group and the galaxy-group separation. Galaxies are then assigned to the group for which they have the highest probability. Following this assignment, the properties of the group are evaluated again and the whole process repeated iteratively until there is no change in the group memberships. Yang et al. (2005) argue that this approach is more successful than the conventional FOF approach and is capable of assigning galaxies in a common dark matter halo to a single group. An alternative method, used by Gerke et al. (2005, 2012) to identify groups in the DEEP2 Galaxy Redshift Survey, is the Voronoi-Delauney method, which Marinoni et al. (2002) showed to be successful at reproducing the distribution of groups velocity dispersions for a DEEP2-like galaxy survey. The Voroni-Delauney method works by dividing space up into polyhedral subvolumes such that there is only one galaxy in each. The volume of each subvolume will therefore act as proxy for the galaxy number density, with highly clustered galaxies occupying smaller subvolumes. Groups are identified by placing progressively larger search volumes and linking together galaxies in neighbouring subvolumes. Murphy et al. (2012) adopt a similar method of using a Voronoi diagram to identify groups in early data from the Panoramic Survey Telescope and Rapid Response System (PanSTARRS) survey. Murphy et al. then assign group members using a follow-up FOF search.

In this chapter we consider the application of a FOF group-finding algorithm to a galaxy survey. Mock catalogues are an essential tool in the calibration of group-finding algorithms as it is impossible to infer the fidelity of the group catalogue analytically. We focus here on the calibration using a synthetic dataset, rather than the application of the group-finder to the real dataset itself. The galaxy survey to which we aim to ultimately apply the group-finder is the Six-degree Field Galaxy Survey (6dFGS Jones et al., 2004, 2009), a shallow (median $z \sim 0.05$), all Southern sky K-band selected galaxy survey. In Section 5.2 we discuss the construction of a 6dFGS mock catalogue and how this catalogue differs from the lightcones described in Chapter 3. In Section 5.3 we introduce the FOF algorithm that we wish to calibrate for use with the 6dFGS. Our goal is to attempt to

calibrate the group-finder by comparing the statistics of the recovered groups with the statistics of the original halos in the mock. We therefore, in Section 5.4, provide details of how we calculate various properties of the groups before, in Section 5.5, discussing further our calibration procedure. In Section 5.6 we present the preliminary results of our parameter search and discuss possible factors affecting the outcome of this search. Finally, we conclude in Section 5.7.

5.2 Constructing a 6dFGS mock galaxy catalogue

Calibration of a group-finding algorithm is essential to ensure that the recovered galaxy groups are a faithful representation of the underlying structures and are not biased due to the choice of group-finding method or parameters. Mock galaxy catalogues are therefore vital tools for performing this calibration (Jing et al., 1998; Eke et al., 2004a). The advantage of using mock catalogues to calibrate the group-finder is that we already know where the groups and clusters are located and which galaxies are members of which groups. With this knowledge we can adjust the algorithm parameters until the recovered groups reproduce the simulated structures as faithfully as possible.

For the calibration to be effective, the mock catalogue adopted must resemble, as closely as possible, the real dataset to which the group-finding algorithm will ultimately be applied. Therefore, we must first consider the specifications of the survey that we wish to mimic.

5.2.1 The 6-degree Field Galaxy Survey

The 6-degree Field Galaxy Survey (6dFGS) is a near-infrared (NIR), K-band selected galaxy survey combining, for the first time, measurements of both spectroscopic redshifts and, for a subset of galaxies, peculiar velocities (Jones et al., 2004, 2009). The survey covers the entire sky visible from the Southern hemisphere down to a galactic latitude limit of $|b| \geq 10^{\circ}$ (approximately $17,000\,\mathrm{deg}^2$) and is complete down to magnitude limits of $(\mathrm{K},\mathrm{J},\mathrm{H}) = (12.65, 13.75, 12.95)$. Of the $136,304$ 6dFGS spectra obtained, ~ 74 per cent were for targets selected from the Extended Source Catalogue (XSC) of the Two Micron All-Sky Survey (2MASS, Jarrett et al., 2000). An additional ~ 9 per cent of optically-selected targets were obtained from SuperCOSMOS[1] (Hambly et al., 2001) with limits of $\mathrm{r_F} \leq 15.60$ and $\mathrm{b_J} \leq 17.65$. As of the final data release (DR3, Jones et al., 2009),

[1]Following the SuperCOSMOS recalibration for the 2dFGRS, (Cole et al., 2005).

the 6dFGS has yielded 110,256 new extragalactic redshifts, with a subset of $\sim 10,000$ galaxies for which peculiar velocities have been obtained. Combined with a further 15,000 measurements from the literature, this gives the 6dFGS a total of 125,071 galaxy redshifts. This is about half the size of the 2-degree Field Galaxy Redshift Survey (2dFGRS, Colless et al., 2001) and ~ 20 per cent of the Sloan Digital Sky Survey (SDSS-II, Abazajian et al., 2009). The 6dFGS is much shallower than the 2dFGRS and the SDSS-II, with a median redshift of $z_{1/2} = 0.053$ compared to $z_{1/2} = 0.11$ for the 2dFGRS and $z_{1/2} = 0.1$ for the main galaxy sample of SDSS-II.

5.2.2 Notable differences with lightcone mocks

Although we have discussed the construction of lightcone mock catalogues in detail in Chapter 3, the mock catalogues constructed here differ in two notable ways.

1. Unlike the lightcone mock catalogues, the mock catalogues we construct for this application are much simpler in the sense that they do not involve interpolating between multiple snapshots from the *Millennium Simulation*, but simply use replications of the redshift zero snapshot of the *Millennium Simulation*. We can make this approximation since the 6dFGS is a fairly shallow survey, with a median redshift of $z_{1/2} \simeq 0.05$, and the K-band luminosity function shows negligible evolution over the volume of the survey.

2. For constructing the lightcone mock catalogues we make use of the post-processed N-body halo merger trees from the *Millennium Simulation* (i.e. consisting of halos that have been post-processed), which means that the final set of post-processed halos does not have a one-to-one correspondence with the FOF halos. Here, however, we use the original FOF halos that have not been post-processed. We choose to use the FOF halos because the group-finder that we wish to calibrate also adopts a FOF algorithm. Our assumption throughout the calibration process is that a single galaxy group is defined as those galaxies that are hosted by the same dark matter halo. Therefore, since we wish to match the properties of the recovered galaxy groups to the properties of the simulation halos, it is reasonable that the two sets of structures are identified using the same algorithm (albeit, in real-space for the halos and redshift-space for the recovered groups).

5.2.3 Assembly of the mock catalogue

As mentioned above, although it is possible to run the GALFORM model directly on the processed halo trees of the *Millennium Simulation*, we do not do this since these halos have been post-processed in the tree construction process and no longer correspond directly to the FOF halos. Our aim is to place GALFORM galaxies directly into the unprocessed FOF halos (as our galaxy group-finding algorithm, which is also a FOF algorithm, should in theory be capable of recovering galaxy groups corresponding to the FOF halos).

To construct the mock catalogue we begin by running the Lagos et al. (2012) version of the GALFORM model on a set of Monte-Carlo generated dark matter halo merger trees, generated following the procedure described in Parkinson et al. (2008). These Monte-Carlo trees have a mass resolution of $5 \times 10^9\, h^{-1}\mathrm{M}_\odot$, slightly better than the mass resolution of the *Millennium Simulation*. In total $\sim$ 2000 trees were generated, with at most 50 halos in each mass bin. The masses of the halos range from the resolution limit up to $\sim 10^{15}\, h^{-1}\mathrm{M}_\odot$.

Next we sample the dark matter particles of the FOF halos in the *Millennium Simulation* (Springel et al., 2005). For each FOF halo, we calculate and store the position and velocity of the centre-of-mass of the halo. We extract the positions and velocities of 20 per cent of the remaining particles in the halo (which we constrain to be between a minimum of 5 and maximum of 10000 particles). For each FOF halo, we search the Monte-Carlo trees for a halo that is closest in mass to the original FOF halo. (If multiple Monte-Carlo halos provide a satisfactory match, then a halo is chosen at random). Once a Monte-Carlo halo has been identified, the central galaxy in the Monte-Carlo halo is placed at the centre-of-mass of the FOF halo, whilst the satellite galaxies are each randomly assigned one of the remaining particle positions.

At this point, so that the mock catalogue is constructed from a population of galaxies with the same luminosity function, we replace the absolute K-band magnitudes output by GALFORM with a set of magnitudes, generated by Monte-Carlo sampling of the 6dFGS K-band luminosity function. We do, however, retain the luminosity ranking of the galaxies as predicted by GALFORM. As can be seen in Figure 5.1, the resultant luminosity function of the simulated galaxies in the cube is now identical to the real 6dFGS luminosity function.

A cosmological volume of sufficient size to contain the entire 6dFGS was generated by stacking replicated copies of the $z = 0$ snapshot volume. We shall refer to the assembled set of stacked simulation volumes as the "supercube". Periodic replication is necessary

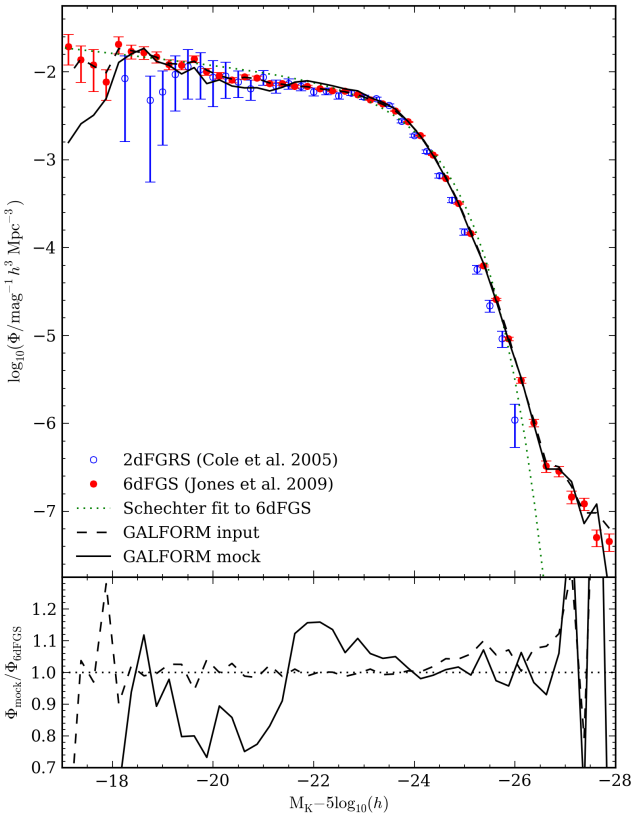

Figure 5.1: Top: K-band galaxy luminosity function for the final 6dFGS catalogue containing approximately 75000 galaxies between $8.65 \leqslant \mathrm{K_{Vega}} \leqslant 12.65$ (red,filled circles). For comparison the K-band luminosity function from the 2dFGRS (Cole et al., 2005) is also shown (blue,empty circles). The dashed line shows the luminosity function of the rescaled `GALFORM` output. The solid line shows the $1/V_{\mathrm{max}}$ luminosity function of the `GALFORM` mock catalogue, with the same flux limits as the 6dFGS. The dotted, green line shows the Schechter fit to the 6dFGS from Jones et al. (2006). Bottom: The ratio of the rescaled `GALFORM` luminosity function to that of the 6dFGS (dashed line) and the ratio of the luminosity function of the `GALFORM` mock catalogue to that of the 6dFGS (solid line).

because, given the cosmology used by the Millennium Simulation, the sides of the simulation box ($500\,h^{-1}$Mpc) correspond to a redshift of $z = 0.17$. In order to contain the entire 6dFGS a minimum of 4 copies of the simulation volume is required.

An observer is placed into the "supercube" with a randomly assigned position and orientation, defined as line-of-sight vector relative to the Cartesian axes of the "supercube", $(\hat{\mathbf{X}}, \hat{\mathbf{Y}}, \hat{\mathbf{Z}})$. The positions of the galaxies in each copy of the simulation cube are rescaled relative to the observer, located at the origin of the supercube. An observer co-ordinate system, $(\hat{\mathbf{X}}', \hat{\mathbf{Y}}', \hat{\mathbf{Z}}')$, is set up such that $\hat{\mathbf{Z}}'$ is along the central line-of-sight of the observer (pointing towards the zenith on the mock southern sky). $\hat{\mathbf{X}}'$ is then defined as orthogonal to both $\hat{\mathbf{Z}}'$ and $\hat{\mathbf{X}}$ and, finally, $\hat{\mathbf{Y}}'$ is defined to be orthogonal to both $\hat{\mathbf{X}}'$ and $\hat{\mathbf{Z}}'$.

By converting the positions of a galaxy into the $(\hat{\mathbf{X}}', \hat{\mathbf{Y}}', \hat{\mathbf{Z}}')$ co-ordinate system, the right ascension, α, and declination, δ, of the galaxy can be calculated. If $\tilde{\mathbf{r}}'(X', Y', Z')$ is the galaxy position in the $(\hat{\mathbf{X}}', \hat{\mathbf{Y}}', \hat{\mathbf{Z}}')$, co-ordinate system, then the right ascension of the galaxy is found by projecting $\tilde{\mathbf{r}}'$ onto the $(\hat{\mathbf{X}}' - \hat{\mathbf{Y}}')$ plane and taking the dot product with $\hat{\mathbf{X}}'$. Similarly, the declination can be found by again projecting $\tilde{\mathbf{r}}'$ onto the $(\hat{\mathbf{X}}' - \hat{\mathbf{Y}}')$ plane and now taking the dot product with the original position vector, $\tilde{\mathbf{r}}'$.

A value for the galactic latitude, b, of a galaxy can then be determined using,

$$\begin{aligned} \sin(b) = \quad & \sin(\delta)\cos(62.6^\circ) \\ & - \cos(\delta)\sin(\alpha - 282.25^\circ)\sin(62.6^\circ). \end{aligned} \tag{5.1}$$

To match the solid angle of the 6dFGS, all galaxies with $\delta > 0^\circ$ and $|b| < 10^\circ$ are discarded.

For all intrinsic galaxy properties (apart from the K-band magnitude), we assume the $z = 0$ values output by `GALFORM`. We choose to simply replicate the $z = 0$ snapshot, rather than interpolate between outputs from snapshots at different epochs, because of the relatively small evolution that is seen in the K-band evolution over the redshift extent of the 6dFGS.

The co-moving separation between the observer and a galaxy is used to calculate a cosmological redshift, $z_{\rm cos}$. Combining this with the local peculiar motion of the galaxy, an observed redshift, $z_{\rm obs}$, can be calculated,

$$z_{\rm obs} = (1 + z_{\rm cos})\left(1 + \frac{v_{\rm r}}{c}\right) - 1 \tag{5.2}$$

where $v_{\rm r}$ is the radial component of the peculiar velocity vector of the galaxy. As Figure 5.2 shows, with the peculiar velocities included, redshift-space effects, such as the 'Fingers of God', are now clearly visible in the mock catalogue.

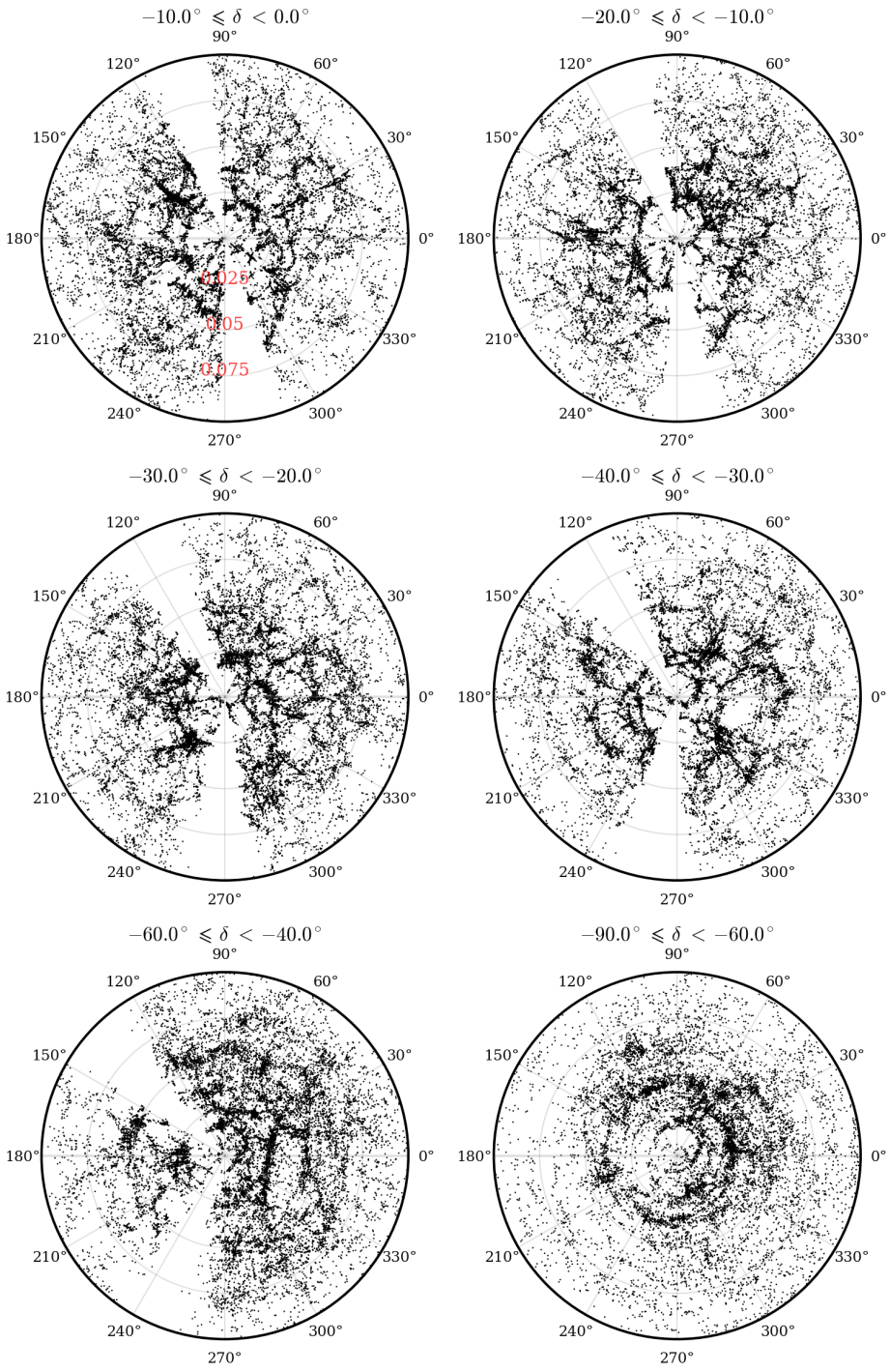

Figure 5.2: Polar plots showing the redshift-space galaxy distribution and large-scale structure within the 6dFGS mock in different declination slices, as labelled above each panel. The plots extend out to redshift $z = 0.1$. Redshift intervals of $\Delta z = 0.025$ are labelled in the top left plot. Peculiar velocities have been included in the redshifts of the mock galaxies.

We then calculate the apparent magnitude of each galaxy, in the band X, using,

$$\begin{aligned} m_{\rm X} = \quad & M_{\rm X} + 5\log_{10}\left(\frac{d_L(z_{\rm cos})}{10{\rm pc}}\right) \\ & -2.5\log_{10}\left(1+z_{\rm cos}\right) + k_{\rm X}(z_{\rm cos}), \end{aligned} \tag{5.3}$$

where $d_L(z_{\rm cos})$ is the luminosity distance out to $z_{\rm cos}$ and $k_{\rm X}(z_{\rm cos})$ is the k-correction required for a galaxy, located at $z_{\rm cos}$, observed in the X-band. Note that the k-corrections used here differ from those used in constructing the lightcone mock catalogues (described in Chapter 3). Instead of applying a frequency shift to the galaxy SED predicted by `GALFORM`, these k-corrections have been calculated by Poggianti (1997) using a separate evolutionary synthesis model, whose results have been compared to the corrections for nearby galaxies.

5.2.4 Post-processing of the mock

At this stage, the mock catalogue resembles an idealized 6dFGS. In reality the survey will suffer from incompleteness due to fibre collisions and observational limitations. This is incorporated in the survey completeness mask, which describes the total redshift completeness of the survey, $T(\theta, {\rm K})$, as a function of sky position, $\theta = (\alpha, \delta)$, and apparent, K-band magnitude,

$$T(\theta, {\rm K}) = R(\theta)C({\rm K})\frac{\int_{\rm K_{bright}}^{\rm K_{faint}} \overline{N({\rm K})}{\rm dK}}{\int_{\rm K_{bright}}^{\rm K_{faint}} C({\rm K})\overline{N({\rm K})}{\rm dK}}, \tag{5.4}$$

where $R(\theta)$ is the completeness as a function of sky position θ (equal to the fraction of targets in the parent input catalogue for which extragalactic redshifts were obtained), $\overline{N({\rm K})}$ is the number of galaxies per unit area (averaged over the whole survey) and $C({\rm K})$ is a function describing the K-band magnitude dependence of the radial completeness. A full description of the completeness mask can be found in Jones et al. (2006). In the 6dFGS, only galaxies for which $T(\theta, {\rm K}) \leqslant 0.6$ are selected for the final galaxy dataset.

For the work in this thesis we do not apply the completeness mask and instead adopt $T(\theta, {\rm K}) = 1$ for all galaxies in the mock. We aim, however, to apply the mask to the mock catalogue in future work.

5.3 The group-finding algorithm

In this work we adopt a group-finding algorithm that uses the popular *friends-of-friends* (FOF, Davis et al., 1985) algorithm. In real space, the FOF algorithm identifies groups

of objects by searching within a spherical volume around each object. This volume is described using a *linking length*, ℓ, which is defined as,

$$\ell = bn^{-1/3}, \tag{5.5}$$

where b is the *linking parameter*, usually set to a fraction of the mean inter-object separation[2] and n is the mean number density of the objects of interest. Any object that is located within the linking volume of another object is defined as being a "friend" of that object. The FOF algorithm then searches recursively through the linking volumes of all of the "friends" to assemble the complete set of associated objects. The simplicity of the algorithm and the lack of assumptions regarding the geometry of halos makes FOF an attractive choice for a group-finder (Giuricin et al., 2000). However, one notable disadvantage is that the algorithm may link together separate structures that are not gravitationally bound. In simulations, this is often corrected by post-processing the FOF structures using a (sub-)halo-finder, such as `SUBFIND` (Springel et al., 2001).

In redshift space, identifying which group a galaxy belongs to is much more difficult. Firstly, the peculiar velocities of galaxies will greatly distort the appearance of structures, particularly along the line-of-sight direction. As a result, structures become smeared out and overlap. Secondly, the flux-limited selection used in galaxy surveys means that groups and clusters at higher redshifts are sampled much more sparsely, as only the brightest members are visible. A successful group-finding algorithm must be able to compensate for such effects.

Application of the FOF algorithm to galaxy surveys therefore typically requires the use of two linking lengths to define the linking volume: one parallel to the line of sight, $\ell_\parallel$, and one perpendicular to the line of sight, $\ell_\perp$. Having $\ell_\parallel \geqslant \ell_\perp$ accounts for the smearing of structures along the line-of-sight, while scaling the linking lengths with the number density of galaxies as a function of redshift, $n(z)$, accounts for the flux-limited selection function. As such, galaxy groups with similar over-density can be identified at all redshifts within the survey volume, which helps prevent the inferred properties of groups (e.g. velocity dispersion, halo mass, projected size) from becoming biased due to the galaxy sampling rate. The challenge is therefore to find optimal values for the constants of proportionality applied in the above relations for the linking lengths.

[2]Application to N-body simulations has shown $b = 0.2$ to be a suitable choice for identifying structures with a density contrast of $\Delta_c \sim 200$ and predicts a halo mass function that is universal, with a shape that is independent of redshift (Jenkins et al., 2001).

The group-finding algorithm that we adopt is the group-finding algorithm that was presented by Eke et al. (2004a) and applied to the 2dFGRS to construct the 2dFGRS Percolation-Inferred Galaxy Group (2PIGG) Catalogue. The algorithm is based upon the approach adopted by Huchra & Geller (1982), whereby two linking lengths, $\ell_\perp$ and $\ell_\parallel$, are used to describe a cylindrical linking volume. The volume is scaled with $n(z,\theta)$, the galaxy number density as a function of redshift and position on the sky, using two free parameters, $b_\perp$ and R, defined by,

$$\ell_\perp = b_\perp n(z,\theta)^{-1/3} \tag{5.6}$$

and

$$\ell_\parallel = R\, b_\perp n(z,\theta)^{-1/3}, \tag{5.7}$$

where $R = b_\parallel / b_\perp$ is the aspect ratio of the cylinder, given by ratio of the linking parameters in the directions parallel and perpendicular to the line-of-sight.

Eke et al. (2004a) calculate the galaxy number density as function of redshift by integrating over the product of the b_J-band galaxy luminosity function and the 2dFGRS selection function, which describes the survey completeness as a function of redshift and b_J galaxy magnitude. We adopt the same approach for the 6dFGS and integrate over the product of the K-band luminosity function, $\Phi(\mathrm{K}, z)$, and the 6dFGS selection function, which is defined by $T_\mathrm{F}(\theta, \mathrm{K})$, the total redshift completeness as a function of sky position, θ, and K-band apparent magnitude.[3] Therefore the galaxy number density is determined by,

$$n(z,\theta) = \int_{\mathrm{K_{bright}}}^{\mathrm{K_{faint}}} \Phi(\mathrm{K}, z)\, T_\mathrm{F}(\theta, \mathrm{K})\, \mathrm{dK}, \tag{5.8}$$

where $\mathrm{K_{bright}}$ and $\mathrm{K_{faint}}$ are the bright and faint apparent K-band limits of the 6dFGS respectively. The number density of the 6dFGS and the `GALFORM` mock catalogue are shown in Figure 5.3.

As a precaution, Eke et al. (2004a) also defined a maximum limit for the linking lengths such that in regions where the galaxy population is sparsely sampled (mainly at high redshift), the scaling relations in Eq. (5.6) and Eq. (5.7), do not inflate the linking volume beyond the size of realistic, dissipationless bound structures. Therefore, $\ell_\perp$, is determined according to,

$$\ell_\perp = \min[L_{\perp,\mathrm{max}}(1+z), b_\perp n(z,\theta)^{-1/3}], \tag{5.9}$$

[3] Further details regarding the calculation of $T_\mathrm{F}(\theta, \mathrm{K})$ can be found in Section 5.2.3.

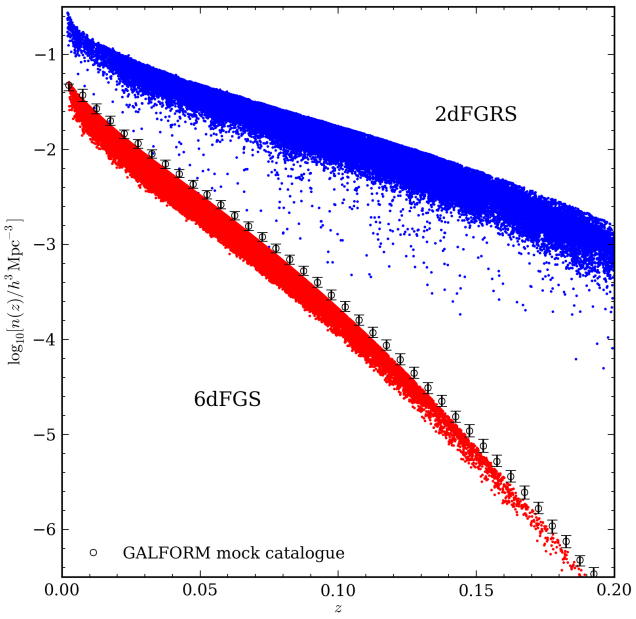

Figure 5.3: The galaxy number density as a function of redshift for the 6dFGS (red) and for $\sim$ 80000 galaxies in the NGP region of the 2dFGRS (blue). The empty circles, with error bars show the median and 10 and 90 percentiles for the number density of the `GALFORM` mock catalogue.

where $L_{\perp,\mathrm{max}}$ is a parameter that sets the maximum linking length permitted in the perpendicular direction.[4]

Marinoni et al. (2002) argue that having a search cylinder whose radius is fixed, is insensitive to local variations in the local density of targets and that the assignment of cluster members will be dependent on the length scale chosen. Eke et al. also note that variations in the local density and the halo concentration will lead to a small systematic bias in the properties of the recovered groups. To correct for this the 2PIGG group-finder is able to fine-tune the linking parameters by running the galaxy sample through the FOF algorithm twice. In the first instance the cylindrical search volume is given a fixed co-moving radius of $1.5\,h^{-1}$Mpc and aspect ratio R. The galaxy number density within the cylinder, relative to the background, is calculated to provide an estimate of the density contrast, Δ, for that structure. This is used to adjust the linking lengths and remove the systematic bias in the second run. As a result the linking lengths are modified according to,

$$b'_{\perp} = b_{\perp}\left(\frac{\Delta}{\Delta_{\mathrm{fit}}}\right)^{\epsilon_b} \tag{5.10}$$

[4]Note that the maximum linking length $L_{\perp,\mathrm{max}}$ is defined in physical units whereas, since they are required to increase with redshift, the linking lengths $l_{\parallel}$ and $l_{\perp}$ are defined in a co-moving length units.

and

$$R' = R\left(\frac{\Delta}{\Delta_{\rm fit}}\right)^{\epsilon_R}, \tag{5.11}$$

where $\Delta_{\rm fit}$, ϵ_b and ϵ_R are parameters to be optimised. In this work we do not attempt to use this fine-tuning feature of the algorithm and so fix $\varepsilon_b = 0$ and $\varepsilon_R = 0$.

Further details regarding the 2PIGG group-finding algorithm can be found in Eke et al. (2004a).

5.4 Galaxy group properties

In this Section we review selected group properties, describing how they are estimated as well as possible systematics.

5.4.1 Richness

The observed number of members in a galaxy group, also known as the observed richness, $N_{\rm obs}$, is one of the simplest, directly measurable properties of a group and can be used as an approximate indicator of the "environment" in which a galaxy resides, by giving an impression of the size of the host group. However, this property is simply the number of member galaxies that are visible above a flux limit and so will be biased by the selection function of the galaxy survey. As a result, groups at different redshifts, with the same observed richness, may, in reality, have vastly different numbers of members.

To compare the richness of galaxy groups (at different redshifts or even, simply, at different positions on the sky) it is preferable to compare the number of members a galaxy group has above a fixed absolute magnitude. One can attempt to account for all of the galaxy members that are fainter than the flux limit, and therefore estimate the true richness, $N_{\rm true}$ of a group, by integrating over the galaxy luminosity function, $\Phi(L)$, assuming that it is the same for galaxies in groups of different mass. The true richness of a group above a fixed luminosity limit, $L_{\rm lim}$, can therefore be estimated from,

$$N_{\rm true} = \frac{\int_{L_{\rm lim}}^{\infty} \Phi(L){\rm d}L}{\int_{L_z}^{\infty} \Phi(L){\rm d}L} N_{\rm obs}, \tag{5.12}$$

where $N_{\rm obs}$ is the observed richness of the group, equal to the number of group members visible above the luminosity limit, L_z, at the redshift of the group.

There is growing evidence that the luminosity function of galaxies varies with the mass of the halo hosting the galaxies. A change in the faint-end slope of the luminosity function with galaxy environment was predicted by Benson et al. (2003) and has since

been observed (e.g. Balogh et al., 2001; De Propris et al., 2003; Blanton et al., 2005b; Baldry et al., 2008), suggesting an increased number of faint galaxies in galaxy clusters.

We note, therefore, that if the above correction is carried out using a luminosity function that is universal over all environments, then a systematic error in the number of galaxies may be introduced, such as in the correction leading to an under-estimate of the number of faint galaxies in a galaxy group. Ideally, therefore, one would wish to integrate over an environment-dependent luminosity function. Doing this for a set of real galaxies, for which estimates of the masses of their host halos are not immediately available, would mean that an additional measure of environment would be required.

In this work, we do not attempt to correct the richness of groups and simply work with their observed richness. In any instances where the estimate of the true richness is required, we assume a universal luminosity function that is independent of the environment of a galaxy.

5.4.2 Luminosity

Several authors have argued that the total luminosity of a galaxy group is a more robust indicator of its mass than the richness or the velocity dispersion, the latter of which may be noisy for small groups (Eke et al., 2004b; Miller et al., 2005; Yang et al., 2005, 2007). The luminosity of a group will be dominated by the brightest few galaxies in the group (the luminosity of which we are assuming correlates well with halo mass), which will typically be successfully recovered by any group-finding algorithm.

Summing up the luminosity of all of the recovered members of a group provides an estimate of the observed luminosity, $L_{\rm obs}$, of a group. The luminosity, in units of solar luminosity, of a galaxy with absolute magnitude, $M_{\rm X}$, is,

$$L_{\rm X} = 10^{0.4\left(M_{\rm X,\odot} - M_{\rm X}\right)} \tag{5.13}$$

where X is the particular waveband of interest (e.g. K or $\rm b_J$) and $M_{\rm X,\odot}$ is the absolute magnitude of the sun in that waveband. For K-band selected galaxies in the 6dFGS we adopt $M_{\rm K,\odot} = 3.28$ (Binney & Merrifield, 1998).

The observed luminosity, however, does not account for the faintest group members, which are fainter than the flux limit of the survey. The total luminosity of a group, $L_{\rm tot}$, can be estimated, similar to the group richness, by integrating over the galaxy luminosity function,

$$L_{\rm tot} = \frac{\int_0^\infty L\Phi(L){\rm d}L}{\int_{L_z}^\infty L\Phi(L){\rm d}L} L_{\rm obs}, \tag{5.14}$$

where $L_{\rm obs}$ is the total luminosity of all of the galaxy members that are visible in the survey, i.e. with a luminosity, $L \geqslant L_z$, greater than the limiting luminosity at the redshift of the group, L_z. If we assume that the luminosity function can be described by a Schechter (1976) function,

$$\Phi(L)\mathrm{d}L = \Phi^{\star}\left(\frac{L}{L^{\star}}\right)^{\alpha}\exp\left(-\frac{L}{L^{\star}}\right)\mathrm{d}\left(\frac{L}{L^{\star}}\right), \tag{5.15}$$

where α, $L^{\star}$ and $\Phi_{\star}$ are the usual Schechter parameters, then the above integral can be expressed by

$$L_{\rm tot} = \frac{\Gamma(\alpha+2)}{\Gamma\left(\alpha+2, L_z/L^{\star}\right)}L_{\rm obs}, \tag{5.16}$$

where $\Gamma(x)$ and $\Gamma(x,y)$ are the complete and incomplete Gamma functions. In a similar way, the total luminosity of a group above a specified but non-zero limiting luminosity, $L_{\rm lim}$, can also be expressed using the relation,

$$L_{\rm tot} = \frac{\int_{L_{\rm lim}}^{\infty} L\Phi(L)\mathrm{d}L}{\int_{L_z}^{\infty} L\Phi(L)\mathrm{d}L}L_{\rm obs} = \frac{\Gamma\left(\alpha+2, L_{\rm lim}/L^{\star}\right)}{\Gamma\left(\alpha+2, L_z/L^{\star}\right)}L_{\rm obs}. \tag{5.17}$$

Note that, as with the richness, any variation in the luminosity function with galaxy environment could lead to biased estimates of the group luminosities, since we have assumed an environment-independent luminosity function.

5.4.3 Velocity Dispersion

The velocity dispersion, σ, is arguably one of the most important properties of a galaxy group and is necessary for estimating the dynamical mass of the group.

Classically, the velocity dispersion is regarded as the standard deviation of the velocity distribution (assumed to be Gaussian) of the group members (Danese et al., 1980). For rich clusters this may be acceptable but in smaller groups, where contamination from interlopers can cause more significant non-Gaussian deviations, the mean and standard deviation are no longer reliable.

To calculate the velocity dispersion we adopt the *gapper* estimator, which has been shown to be robust for small-number statistics (Beers et al., 1990). The estimator works by taking an ordered set of N velocities, v_i, v_{i+1}, ..., v_N, and defining gaps, g_i, between the velocities as

$$g_i = v_{i+1} - v_i, \qquad i = 1, 2, ..., N-1. \tag{5.18}$$

By weighting the gaps

$$w_i = i(N-i), \tag{5.19}$$

the velocity dispersion can be estimated as

$$\sigma_{\rm gap} = \frac{\sqrt{\pi}}{N(N-1)} \sum_{i=1}^{N-1} w_i g_i \tag{5.20}$$

Following Eke et al. (2004a), we define the line-of-sight velocity dispersion, $\sigma_{\parallel}$, of a group to be

$$\sigma_{\parallel} = \sqrt{\max\left(0, \frac{\mathrm{N}\sigma_{\rm gap}^2}{\mathrm{N}-1} - \sigma_{\rm err}^2\right)}, \tag{5.21}$$

where $\sigma_{\rm err}$ is the redshift measurement error. Note that before calculating the velocity dispersion, we use 3σ clipping to remove extreme outliers. If we assume that the group is spherically symmetric, then the line-of-sight velocity dispersion is related to the 3D velocity dispersion, σ, by

$$\sigma_{\parallel} = \sigma/\sqrt{3}, \tag{5.22}$$

(Limber & Mathews, 1960; Bahcall & Tremaine, 1981; Heisler et al., 1985; Merchán & Zandivarez, 2002).

5.4.4 Radius

Before calculating the radius of a group, one must have an estimate for the location of the group centre. In this work we locate the centres of groups following the method used by Eke et al. (2004a), who determined the projected centre through an iterative process to calculate the arithmetic weighted mean sky position of the N member galaxies. After each iteration, the weighted projected separation between the group centre and each remaining member is calculated and the most distant galaxy is ejected. This is repeated until only two galaxies remain. The projected centre of the group is then placed on the galaxy with the greatest weight (or flux if the two galaxies have the same weight). The sky position and redshift of this galaxy are used as the sky position and redshift, $z_{\rm grp}$, of the group centre.

The physical size of a group is commonly estimated using the projected virial radius, $r_{\rm p,vir}$, which is defined,

$$r_{\rm p,vir} = \frac{N(N-1)}{\sum_{i>j} r_{ij}^{-1}}, \tag{5.23}$$

where N is the number of group members and r_{ij} is the projected separation between galaxies i and j, (Heisler et al., 1985; Merchán & Zandivarez, 2002; Crook et al., 2007). Assuming spherical symmetry, this can be related to the 3-dimensional virial radius, $r_{\rm vir}$ of the group by,

$$r_{\rm vir} = \frac{\pi}{2} r_{\rm p,vir}, \tag{5.24}$$

(Limber & Mathews, 1960; Bahcall & Tremaine, 1981).

Eke et al. (2004a) instead use the root mean square (rms) projected radius, $r_{\rm p}$, of the group, calculated as follows,

$$r_p = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(w_i r_i)^2}. \tag{5.25}$$

where r_i is the projected separation of galaxy i from the group centre and w_i is a weight equal to the inverse of the total redshift completeness around that galaxy, i.e. $w_i = 1/T_{\rm F}(\theta_i, {\rm K}_i)$. Note that in this work we are not applying any completeness mask and are thus assuming $w_i = 1$. Since one of the galaxy members is assumed to be located at the centre of the group, the summation in Eq. (5.25) is carried out over the remaining $N-1$ group members.

5.4.5 Mass

Although in this work we do not calculate the mass of any of the recovered groups, we plan to calculate the masses of the 6dFGS groups in future work and so briefly review how the mass of the groups can be estimated.

A dynamical estimate of the mass of the dark matter halo that hosts a galaxy group can be made by using the velocity dispersion of the group members and applying the virial theorem. The viral mass, $M_{\rm vir}$, of the host halo can therefore be estimated by,

$$M_{\rm vir} = \frac{3\pi}{2}\frac{\sigma_{\parallel}^2 r_{\rm p,vir}}{G}, \tag{5.26}$$

where $r_{\rm p,vir}$ is the projected virial radius, (Bahcall & Tremaine, 1981; Heisler et al., 1985).

In their application of the group-finding algorithm to the 2dFGRS, Eke et al. (2004a) adopted a similar mass estimator, where the mass, $M_{\rm h}$, of the halo is calculated according to,

$$M_{\rm h} = A\frac{\sigma_{\parallel}^2 r_p}{G}, \tag{5.27}$$

where A is a constant of proportionality, which Eke et al. fix by ensuring that the value chosen leads to unbiased median mass estimates. Using a set of 2dFGRS mock catalogues (Cole et al., 1998), they conclude that $A = 5$ is the best choice, which, we note, is very close in value to the coefficient of $3\pi/2$ in Eq. (5.26).

5.5 Calibration procedure

Previous studies have often calibrated group finding algorithms by attempting to perform a one-to-one matching of the recovered groups to the original halos in a mock catalogue (e.g. Eke et al., 2004a; Berlind et al., 2006; Murphy et al., 2012). In order to do this it is necessary to define several statistics to quantify the success of this matching procedure. These include quantities such as the 'completeness', 'purity' and 'fragmentation' of the halos, which can often have specialised definitions that can differ between different authors.

In the application of the 2PIGG group-finder to the 2dFGRS, Eke et al. (2004a) used a series of 2dFGRS mock catalogues (Cole et al., 1998) to calibrate the algorithm for the 2dFGRS. The mock catalogues were used to perform one-to-one matching between the halos and the recovered groups with the aim of identifying the linking length parameters that maximised the number of correct membership assignments and minimised the number of interloper galaxies. Eke et al. constructed several definitions to quantise and measure the success of the matching procedure as a function of the redshift, velocity dispersion and mass of the halos in the mock catalogue. From their calibration, Eke et al. found $b_\perp = 0.13$, $R = 11$ and $L_{\perp,\max} = 2\,h^{-1}$Mpc to be the best choice of parameters for returning a set of 2dFGRS galaxy groups with unbiased velocity dispersions and sizes. Following fine-tuning of the algorithm, Eke et al. selected the parameter values $\Delta_{\rm fit} = 5$, $\epsilon_b = 0.04$ and $\epsilon_R = 0.16$. Note, however, that we are neglecting applying the fine-tuning features of the algorithm for this work.

One of the main criticisms of this approach is that it potentially leads to the group-finding algorithms being over-calibrated to the synthetic dataset adopted. For example, one can perform a one-to-one matching to the halos in a mock constructed using a particular galaxy formation model, but this matching could lead to an entirely different result if an alternative galaxy formation model is assumed. In this work, we therefore attempt to calibrate the chosen group-finding algorithm by instead attempting to match the statistics of the population of dark matter halos in the mock catalogue, which should be more robust against changes in the underlying physics, such as the galaxy formation model assumed.

Berlind et al. (2006) adopted a more statistical approach when calibrating their group finding algorithm for application on the SDSS. They focused on the recovery of the group multiplicity function, which describes the abundance of groups of different richness. Using

a set of four mock catalogues, Berlind et al. determined that for groups with a high richness ($N \gtrsim 10$) the multiplicity function is less sensitive to the choice of linking parameters and that measurements of the group abundance should only really be trusted for groups with high richness. They argued, however, that recovery of an unbiased multiplicity function does not guarantee an unbiased membership assignment and that a one-to-one matching procedure is also necessary. Additionally, Berlind et al. compared the median line-of-sight velocity dispersions and rms projected radii of groups to those of halos, as a function of group richness. They define the halo distribution to be recovered if, for all richnesses considered, the distribution recovered by the groups has median values that lie within 10 per cent of the values for the halos. Their main conclusion from this is that it is extremely difficult to identify a set of linking lengths that provide a group catalogue with unbiased abundances, velocity dispersions and projected sizes simultaneously. Since Berlind et al. wished to consider the occupation distributions of galaxy groups, they sacrificed recovery of unbiased velocity dispersions and adopted linking lengths of (in the notation of Eke et al.) $b_{\perp} = 0.14$ and $R \simeq 5.4$, which, from their calibration, they were confident would lead to improved recovery of group memberships.

To calibrate the group-finding algorithm we consider several statistics of the halo population in the 6dFGS mock catalogue and attempt to see whether we can reach an optimum linking length parameter combination while avoiding undertaking a one-to-one matching assessment. Ultimately, of course, we will have to perform such an assessment to test whether we have been successful. For now, however, we simply wish to search for the optimum set of linking lengths that best recover the distribution of halo properties of our input mock catalogue. We do this by searching through the linking length parameter space with the aim of finding the linking length combination that produces a catalogue of groups with similar statistics to the mock halos. The statistics that we consider are the redshift distribution of the halo/group centres, the distribution of observed group luminosities, the distribution of the line-of-sight velocity dispersions of the halos/groups and the distribution of projected radii of the halos/groups. By comparing the redshift distribution of the recovered groups to that of the halos, we can check to see that we are placing groups in approximately the correct positions. We use the distribution of velocity dispersions and projected sizes to check that we are recovering the correct sizes of the groups in directions parallel and perpendicular to the observers line-of-sight respectively. We compare the observed luminosity distribution as a measure of how well with the groups we are recovering the membership of the halos.

We calculate the properties for the actual halos in the mock from an observational perspective and therefore use the exact same techniques that are used to infer the properties of the recovered groups. This way any differences between the two datasets arise solely due to incorrect membership assignment.

For each of the above listed halo properties, we compare the differential distributions of the halos to the differential distribution of the recovered groups by calculating the χ^2 statistic, which we define as

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(n_{\mathrm{halo},i} - n_{\mathrm{group},i}\right)^2}{n_{\mathrm{halo},i}}, \tag{5.28}$$

where $n_{\mathrm{halo},i}$ and $n_{\mathrm{group},i}$ are the number of halos and groups in each bin, i, respectively and N is the number of bins used to construct the distributions. We calculate the reduced χ^2 by assuming that the system has $N - 1$ degrees of freedom.

Our aim, therefore, is to search for the linking length parameter combinations that minimise the chi-squared statistic for each halo property. By adding the χ^2 for the separate properties, we can also search for the parameter combination that minimises the total χ^2. The advantage of using the χ^2 statistic is that, by dividing by the Poisson error on the halo counts, we are adding greater weight to the peak of the distributions, corresponding to the majority of the halos/groups. One disadvantage is that by binning the data, our choice of bins may introduce a potential systematic bias. We thus need to check that using a different number of bins to construct our distributions does not significantly affect the location of the minimum χ^2.

An alternative method, that avoids binning of the data, is to compare the cumulative distribution of the halos to that of the recovered groups. Each cumulative distribution is constructed by ranking the halos/groups according to the desired property and then using linear interpolation to sample the distribution at N locations. To compare the cumulative distributions, a cost function would need to be constructed. A basic cost function could, for example, simply sum up the absolute differences between the cumulative distributions at the N locations.

5.6 Parameter search results

We begin the parameter search by comparing the distributions of each property obtained when considering all halos and groups in their respective catalogues that have a minimum observed richness of 2, i.e. all halos and recovered groups that host at least 2 galaxies that

are visible above the flux limit. We fix $L_{\perp,\max} = 2\,h^{-1}$Mpc. We correct the luminosities using Eq. (5.14), using a luminosity limit corresponding to an absolute magnitude of $M^\star + 1$, where we have used $M^\star_{\rm K} - 5\log_{10}(h) = -23.83$, as reported by Jones et al. (2006). When calculating the χ^2 statistic for the various properties, we only consider bins containing at least 100 halos or groups, thus giving us an error of 10 per cent.

In Figure 5.4, we show the results of minimising the χ^2 statistic for each halo property separately. The plots in the right-hand column show the value of the χ^2 statistic, minimised for each property separately, throughout the linking-length parameter space, which consists of the linking length perpendicular to the observer's line-of-sight, $b_\perp$, and the axis ratio of the linking length cylinder, R. The symbols plotted on the χ^2 map correspond to the optimum linking lengths obtained when minimising χ^2 for each property in turn (and so are plotted in the same location on each χ^2 map). The plots in the left-hand column show the corresponding distributions for each property at each of these locations in the parameter space and compare them to the distribution for the halos in the mock catalogue.

5.6.1 Degeneracies in the linking length parameter space

A striking feature in the maps of the χ^2 values is the presence of a degeneracy in the values of the linking length parameters that minimise the χ^2 statistic. This degeneracy occurs at small values of $b_\perp \lesssim 0.15$ and appears in the χ^2 map for each of the group properties considered.

At smaller values of R the degeneracy shifts towards larger values of $b_\perp$ with a shape characteristic of a reciprocal function, which is understandable given that, by definition, $R \propto 1/b_\perp$. Such a degeneracy was also found by Crook et al. (2007). This reflects the change in the shape of the linking volume, from needle-like cylinders (with large values of R and small values of $b_\perp$) to pancake-like cylinders (with small values of R and large values of $b_\perp$).

To understand why this degeneracy arises, let us consider an example scenario illustrated in Figure 5.5. First consider fixing $b_\perp$ and allowing R to vary, leading to the two possible linking volumes indicated by the solid and dotted lines. In this situation, if we reduce R we would expect the redshift of the group centre to remain the same. Additionally, if the galaxies identified by the larger linking volume but not by the smaller linking volume are sufficiently faint compared to the galaxies inside the original halo, then it is possible for the observed luminosity of the group to not change significantly. This is

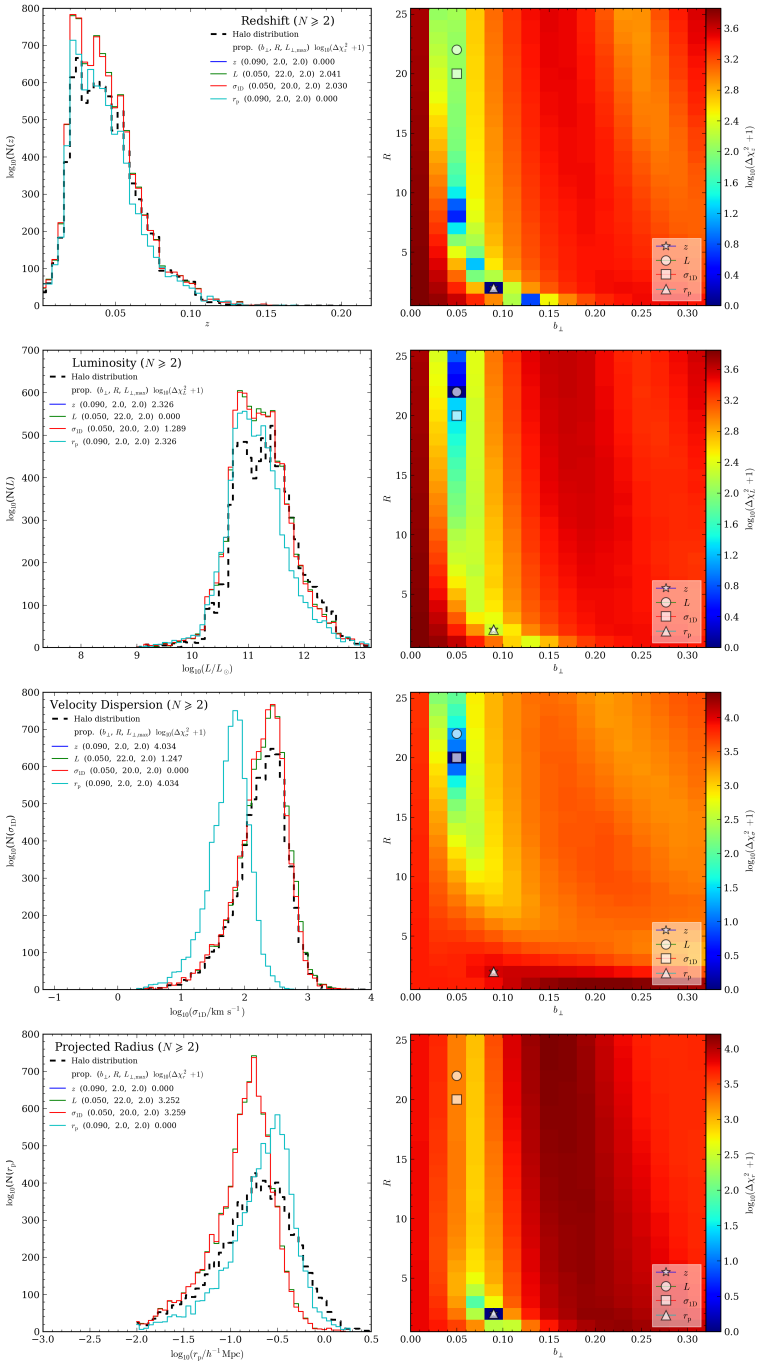

Figure 5.4: Results for minimising the χ^2 statistic for each of the properties: redshift (top row), group luminosity (second row), line-of-sight velocity dispersion (third row) and projected radius (bottom row). The pixel maps show the log difference in the values of χ^2, relative to the minimum χ^2, throughout the linking length parameter space. The various symbols show the locations of the minima for each of the property-specific χ^2. The distributions at these linking length locations are shown in the left-hand plots. Here, only halos and groups that host at least 2 galaxies above the flux limit have been considered. Note that in the χ^2 maps the star and triangle are lying on top of one another and that in the plots of the distributions the blue and cyan lines are lying on top of one another.

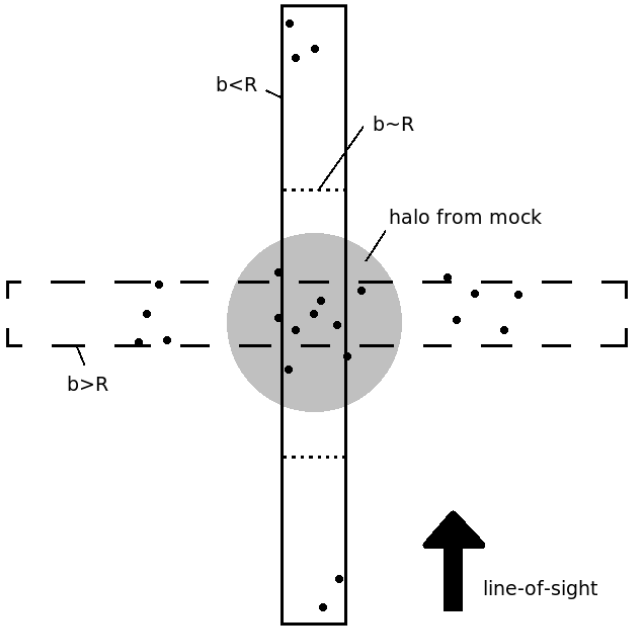

Figure 5.5: Figure illustrating three different linking volumes: a needle-like linking volume ($b_\perp < R$, solid line), a pancake-like linking volume ($b_\perp > R$, dashed line) and a linking volume in between these two extremes ($b_\perp \sim R$, dotted line). In this example situation it is plausible that these three linking volumes could recover a group with a similar central redshift and number of members.

of course a simple demonstration where we have assumed that the number of interloper galaxies that are picked up at redshifts larger than that of the original halo is similar to the number of interloper galaxies that are picked up at lower redshifts.

If we now allow $b_\perp$ to vary and consider the case where $b_\perp$ is very large and R is reduced to a small value, as shown by the dashed line in Figure 5.5, then we can see that it is possible for the central redshift and observed luminosity of the recovered group to be the same as in the other two cases. The degeneracy does not extend as far in the direction of $b_\perp$ because, for very large perpendicular linking lengths, we would begin to merge together many of the groups located at a similar redshift and therefore reduce the counts in that bin of the redshift distribution. This is also why the degeneracy does not extend to extremely large linking volumes, with large values for both $b_\perp$ and R.

The presence of the degeneracy means that there will be many linking length combinations that will recover populations of groups which display a reasonable match to the distribution of halo redshifts and luminosities, as these two properties are not sensitive to the dimensions of the linking volume. Examining the distributions in Figure 5.4 we can see that all of the proposed minima show reasonable agreement with the redshift and

luminosity distributions, though there is slightly better agreement between the redshift and the projected sizes, and between the velocity dispersions and the luminosities. Since the velocity dispersions and projected radii of recovered groups will show the greater dependency on the dimensions of the linking volume, we would expect these two properties to be necessary to help place tighter constraints on the choice of optimum linking lengths. However, if we consider the line-of-sight velocity dispersion and projected radius (bottom two rows in Figure 5.4), then we see that the degeneracy in the χ^2 maps for these properties is weaker than for the redshift and observed luminosity. Although the degeneracy is still present, it is clear that the velocity dispersion favours small values for $b_\perp$ and large values for R, whilst the projected radius favours a larger value for $b_\perp$ and a smaller value for R. Previous authors have noted that it is not possible, using the FOF technique, to simultaneously recover both the line-of-sight velocity dispersion and the projected sizes of groups (e.g. Berlind et al., 2006).

From the χ^2 maps, we can clearly see that there is no region in the linking length parameter space capable of providing a good fit to both the true distributions of velocity dispersions and the projected sizes. We must therefore look for a linking length parameter combination that makes a reasonable compromise. It is instructive therefore to consider the net value for χ^2 obtained when summing up the χ^2 values for each of the properties individually.

In Figure 5.4 a very weak second degeneracy is visible in the χ^2 maps at large values of both $b_\perp$ and R. We stress, however, that this second degeneracy corresponds to a local minimum with χ^2 values that are much larger than the global minimum that we have already identified. The large χ^2 values around this local minimum indicate that the property distributions for groups recovered using such large linking lengths are a much poorer match to the distributions of halo properties.

To understand how this local minimum arises, consider the case where R is kept fixed and $b_\perp$ is allowed to vary. An example is shown in Figure 5.6, where we plot the redshift distributions of catalogues of groups recovered with fixed $R = 20$. By fixing R and increasing $b_\perp$ we maintain the aspect ratio of the cylinder, whilst increasing it in size. Starting from a small value of $b_\perp \sim 0.05$, we see that by increasing the value to $b_\perp = 0.15$ we first increase the number of recovered groups in all redshift bins as more isolated galaxies are linked together. However, by increasing the value of $b_\perp$ further we begin to merge groups of galaxies together, causing the number of recovered groups to decrease.

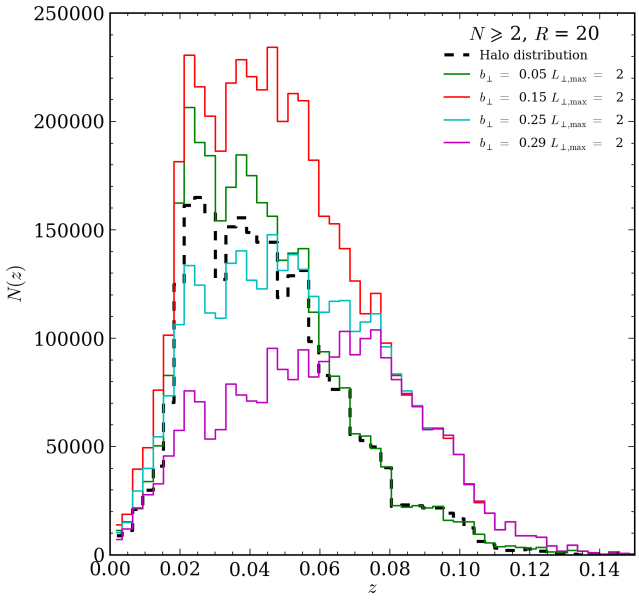

Figure 5.6: Redshift distributions for catalogues of groups recovered using a fixed value of $R = 20$, but with different values for $b_\perp$.

This occurs first at lower redshifts[5]. As the number of groups at lower redshifts is reduced then the peak of the group redshift distribution comes once again into closer agreement with the peak of the halo distribution, i.e. we can see that in Figure 5.6 the peak of the distribution obtained using $b_\perp = 0.25$ (cyan line) is in much better agreement with the peak of the halo distribution compared to the distributions obtained using $b_\perp = 0.15$ (red line) or $b_\perp = 0.3$ (magenta line). We therefore, would expect $b_\perp = 0.25$ to have a χ^2 value that is smaller than that for $b_\perp = 0.15$ or $b_\perp = 0.3$. Note, however, that $b_\perp = 0.25$ has a larger χ^2 than $b_\perp = 0.05$, which is likely due to the poorer disagreement with the high redshift tail of the halo redshift distribution.

5.6.2 Property combinations

In Figure 5.7, we show the net value for χ^2 obtained when summing up the separate property-specific χ^2 in several different combinations. Although the degeneracy is not broken, we can see the effect of introducing different properties into the value for the

[5]One might naively expect this given that the inter-galaxy separation will be smaller at lower redshifts. Recall, however that the linking lengths scale with the number density of galaxies, so we should expect to start merging groups at all redshifts. One possibility is that such behaviour reveals that the scaling with redshift that we have adopted for the linking lengths is not optimal. In this work, however, we are not looking to modify the group-finding algorithm, but simply to calibrate its parameters.

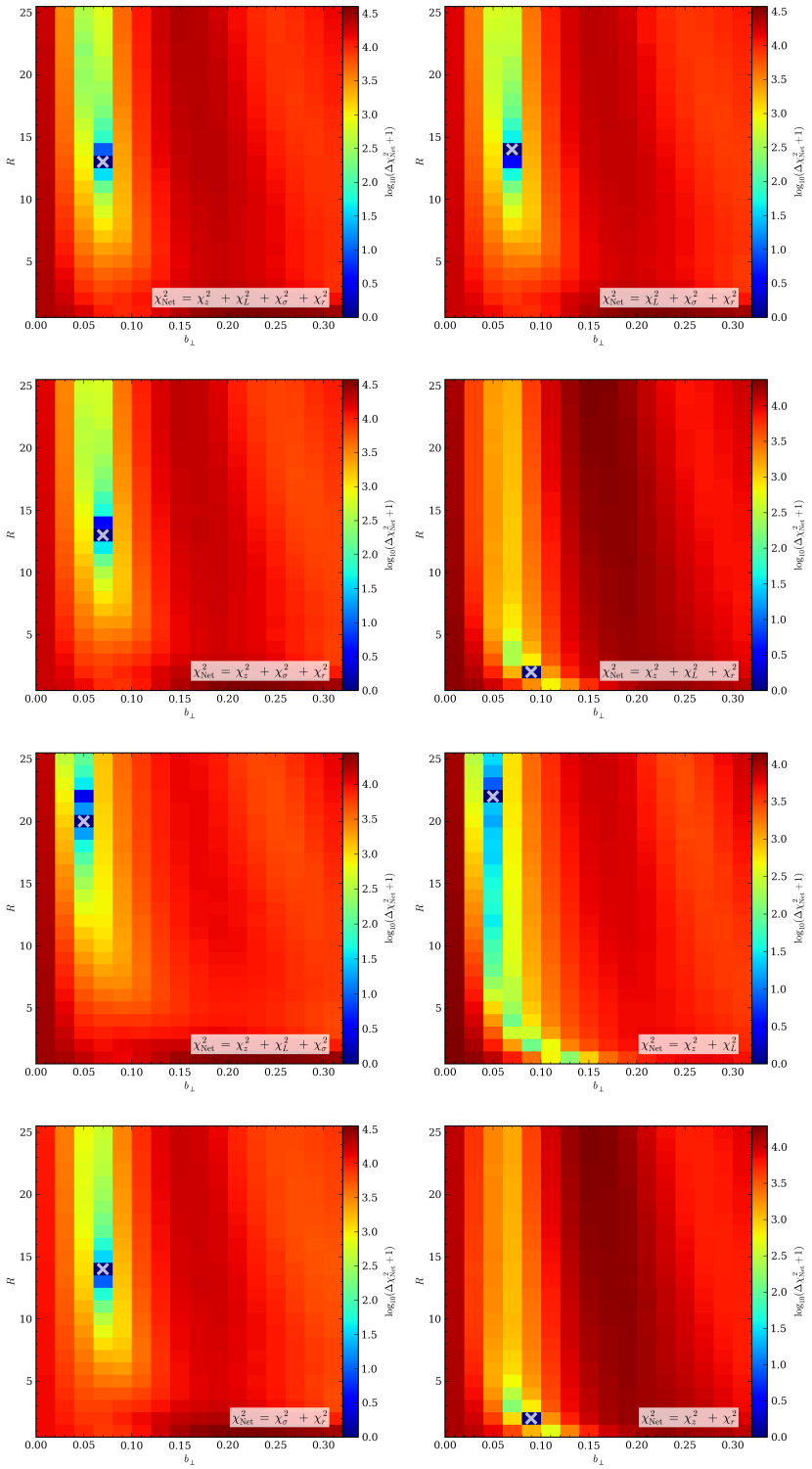

Figure 5.7: Pixel maps showing the locations for the optimum linking lengths after summing the χ^2 for different combinations of properties. The properties used in each case are indicated in the bottom right of each panel. White crosses mark the location of the minimum value of the net χ^2 in each case. Like in Figure5.4, the colour bars indicate the log difference in the values of χ^2 from the minimum value.

net χ^2. In the majority of cases, combining the values for χ^2 predicts an optimum linking length combination of $(b_\perp, R) = (0.07, 14)$. Note that Eke et al. (2004a) recovered $(b_\perp, R) = (0.13, 11)$ following their calibration of the algorithm for use with the 2dFGRS. The smaller value for $b_\perp$ compared to that obtained for the 2dFGRS could be due to the 6dFGS picking up a larger number of small, faint systems at low redshift.

In all instances where one of either the velocity dispersion or projected size has been removed, the optimum linking lengths are biased either too large (if the projected size is removed) or too small (if the velocity dispersion is removed). This supports our expectation that both of these properties are necessary to constrain the geometry of the linking volume. An exception, however, is the prediction of the optimum linking lengths when all four properties are included in the calculation of the net χ^2. In this instance, the linking lengths are biased too large despite both the velocity dispersion and the projected sizes being included in the calculation. Notice, however, that when one of either the redshift or the luminosity is removed, the values $(b_\perp, R) = (0.07, 14)$ are recovered. This could suggest that the luminosity of the group is correlated with its redshift. We would expect a correlation between the observed luminosity of a group and the group redshift, however, since we are using a corrected luminosity down to an absolute luminosity limit, it is not clear why a correlation would be present. Regardless, we will from now on calculate the net χ^2 using both the velocity dispersion and the projected sizes and only one of either the redshift or the luminosity.

In Figure 5.8, we show the property distributions for the various linking lengths identified by crosses in Figure 5.7. An eyeball inspection shows that all of the suggested linking lengths provide reasonable matches to the redshift and luminosity distributions. The parameters $(b_\perp, R) = (0.09, 2)$ are a poor match to the distribution of velocity dispersions, whilst the two linking length combinations that have $b_\perp = 0.05$ are a worse match to the distribution of projected sizes. The linking lengths $(b_\perp, R) = (0.05, 22)$ appear to provide the best match to the distribution of velocity dispersions. The linking lengths $(b_\perp, R) = (0.07, 14)$, which were recovered in the majority of cases in Figure 5.7, appear to provide a good compromise, showing reasonable agreement with all four of the property distributions, though they do over-predict the number of groups at the peak of the distributions, especially for the velocity dispersion and projected radius. In each case, these linking lengths appear to be recovering a greater number of groups than the number of halos in the mock. The greater number of groups with smaller projected sizes suggests that many of the larger groups are still being broken up. Since our χ^2 should be

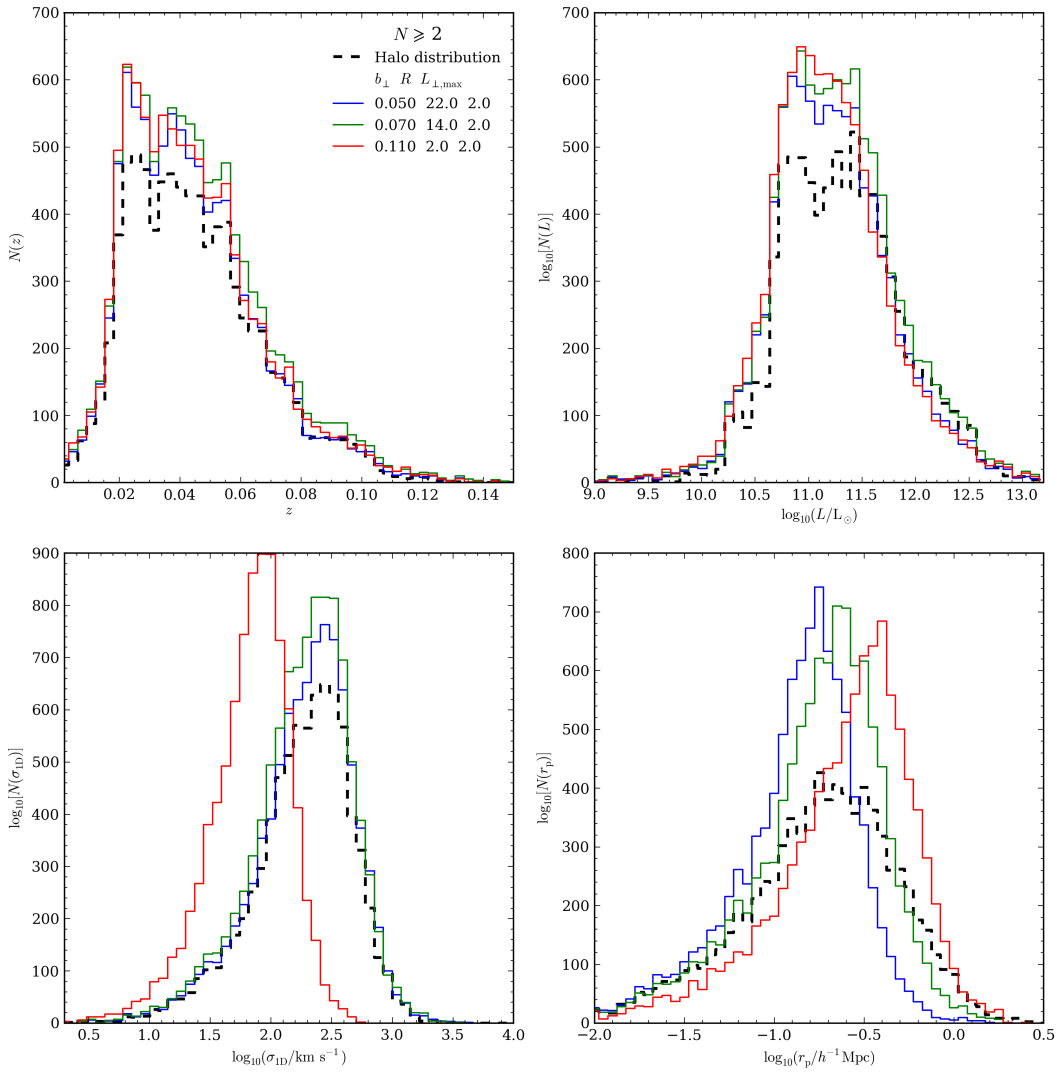

Figure 5.8: Property distributions for the galaxy groups recovered using four different linking length parameter combinations, as shown by the various coloured lines. The top left panel shows the distribution of group redshifts, the top right panel shows the distribution of the observed luminosities of the groups, the bottom left panel shows the distribution of line-of-sight velocity dispersions and the bottom right panel shows the distribution of projected radii of the groups. In each case the black dashed line represents the corresponding distribution for the halos in the input mock catalogue.

biased towards matching the peak of the distributions, it is not clear why these optimum linking lengths appear to give such a poor match at the peak.

Since, in Figure 5.7, the linking lengths $(b_\perp, R) = (0.07, 14)$ are predicted as the location of the minimum net χ^2 value in the majority of the cases considered, and that this parameter combination provides a reasonable match to the distribution of halo properties, we take this parameter combination as the optimum values. There are obviously, however, many factors which could influence this result, which we must investigate, including the effect of the bin size used to construct the property distributions, the range of redshifts over which we perform the comparison (i.e. whether we wish to compare the distribution over all redshifts or, for example, just around the median redshift) and the observed richness of the groups and halos that we use (i.e. whether we match the distribution using all groups or just those for which $N_{\rm min}$ members are visible). In the remaining sections of this chapter, we consider each of these factors in turn.

One factor that would affect the values of the optimum linking lengths is the size of the halos in the mock catalogue. If the linking length used to identify the halos in the simulation was too large then the resulting halos would be too massive, thus leading to our linking lengths being too large. We argue, however, that since the linking length used to identify the halos in the simulation leads to a halo mass function that is universal and has a shape that appears to be independent of redshift, as expected in Press-Schechter theory, then the halos obtained are a reasonable prediction for halos in the real universe.

5.6.3 Changing the binsize

An obvious factor that we need to consider is how the minima in the χ^2 values change if we change the number of bins that we use to construct the distributions of halo and group properties. In Figure 5.9 and Figure 5.10 we adjust the number of bins used from 50 to 20 and to 100 respectively.

If we use 20 bins, our values for the linking lengths do not change appreciably. However, if we switch to using 100 bins, the optimum linking lengths do change. This is understandable since we are increasing the Poisson noise in each bin. Now the optimum linking lengths recovered by matching each property individually are distributed along the degeneracy. If we combine the properties to calculate the net χ^2 values, as shown in Figure 5.11 and Figure 5.12, our original linking lengths of $(b_\perp, R) = (0.07, 14)$ are never recovered, with $(b_\perp, R) = (0.05, 22)$ and $(b_\perp, R) = (0.07, 16)$ being recovered in most cases. We have already chosen to reject $(b_\perp, R) = (0.05, 22)$ since it cannot match both

the velocity dispersion and the projected sizes whilst $(b_\perp, R) = (0.07, 16)$ provide property distributions that are not too dissimilar to those recovered using $(b_\perp, R) = (0.07, 14)$. So, either of $(b_\perp, R) = (0.07, 14)$ or $(b_\perp, R) = (0.07, 16)$ would appear to be acceptable. We are confident, therefore, that the number of bins that we use to construct the halo and groups property distributions is not significantly biasing our recovery of the linking lengths. As such, we shall continue to construct the property distributions using 50 bins.

5.6.4 Changing $L_{\perp,\mathrm{max}}$

One possible change that could be affecting our search for the optimum linking lengths is the value for $L_{\perp,\mathrm{max}}$, which so far we have set to $2\,h^{-1}$Mpc. In Figure 5.13 we again plot the distributions recovered for the various properties if we adopt $L_{\perp,\mathrm{max}} = 3\,h^{-1}$Mpc. Changing the value of $L_{\perp,\mathrm{max}}$ appears to have no appreciable affect on the linking lengths predicted by each property individually. Figure 5.14 shows the net values for χ^2. We find that adopting $L_{\perp,\mathrm{max}} = 3\,h^{-1}$Mpc has no appreciable affect and that $(b_\perp, R) = (0.07, 14)$ still appears to provide the best compromise between the properties.

5.6.5 Tolerance on bin error assumed

So far we have assumed a tolerance of 10 per cent, that is, when calculating the value of χ^2 we have only considered bins that contribute 100 halos/groups or more. We now consider how our results change if we relax this tolerance.

In Figure 5.15 we show the property distributions and χ^2 maps for a tolerance of 30 per cent (i.e. using bins contributing at least 10 groups). Changing the tolerance has no obvious affect on the optimum values for the linking lengths. In Figure 5.16, we show the net χ^2 values obtained using a 30 per cent tolerance. Again, the increase in the tolerance has no appreciable affect on the optimum linking lengths, with $(b_\perp, R) = (0.07, 14)$ still being recovered in most cases.

5.6.6 Effect of redshift range

So far in our parameter search we have attempted to match the distribution of halo properties using groups spanning the whole redshift range of the mock catalogue. Since the majority of the galaxies are found near the median redshift of the 6dFGS ($\bar{z} \simeq 0.05$), we ask how our optimum linking lengths change if restrict our comparison to groups with redshifts around this value. Note that this is not the same as changing the tolerance on

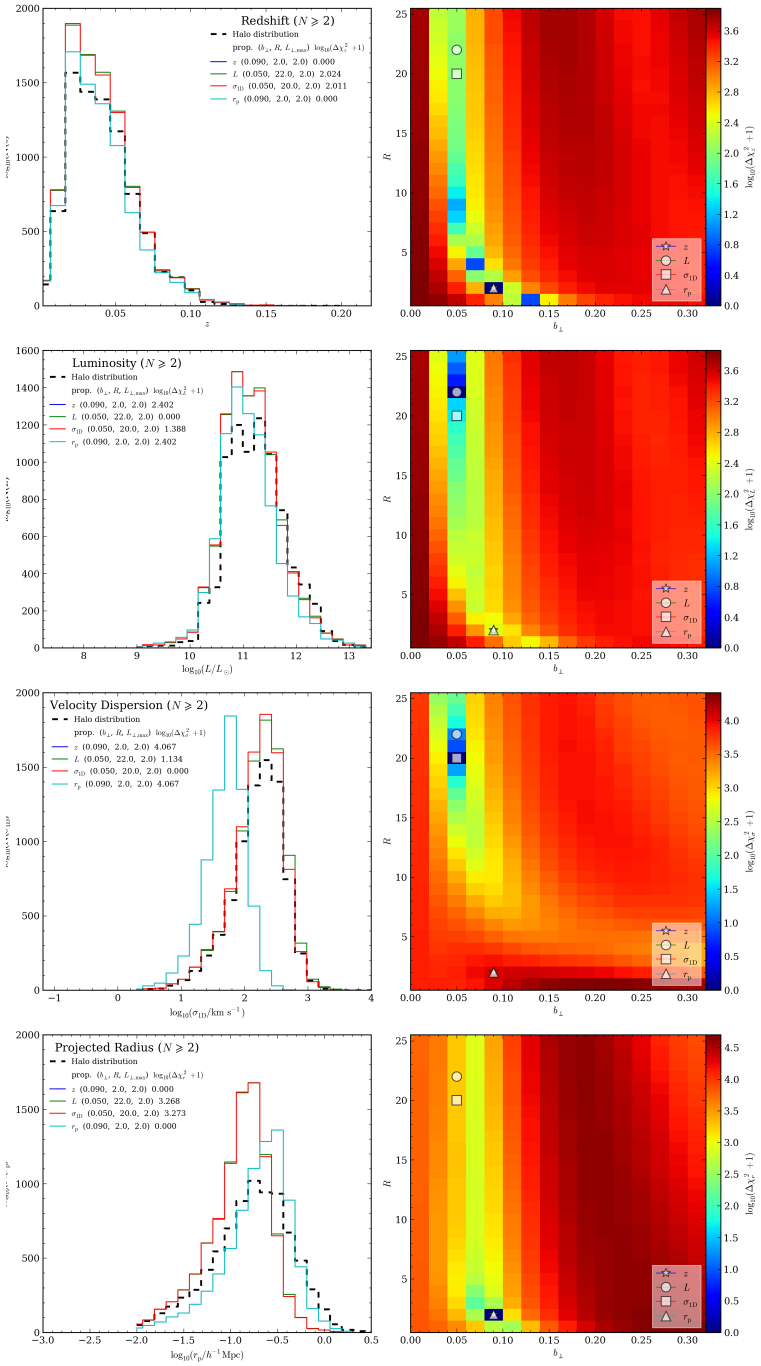

Figure 5.9: Results minimising the property the χ^2 statistic using each of the individual halo and group properties. Here we have constructed the property distributions using 20 bins. The symbols, lines and colour bar have the same meanings as in Figure 5.4. Note that the star and the triangle symbols are lying on top of one another and the cyan line and the blue line are lying on top of one another.

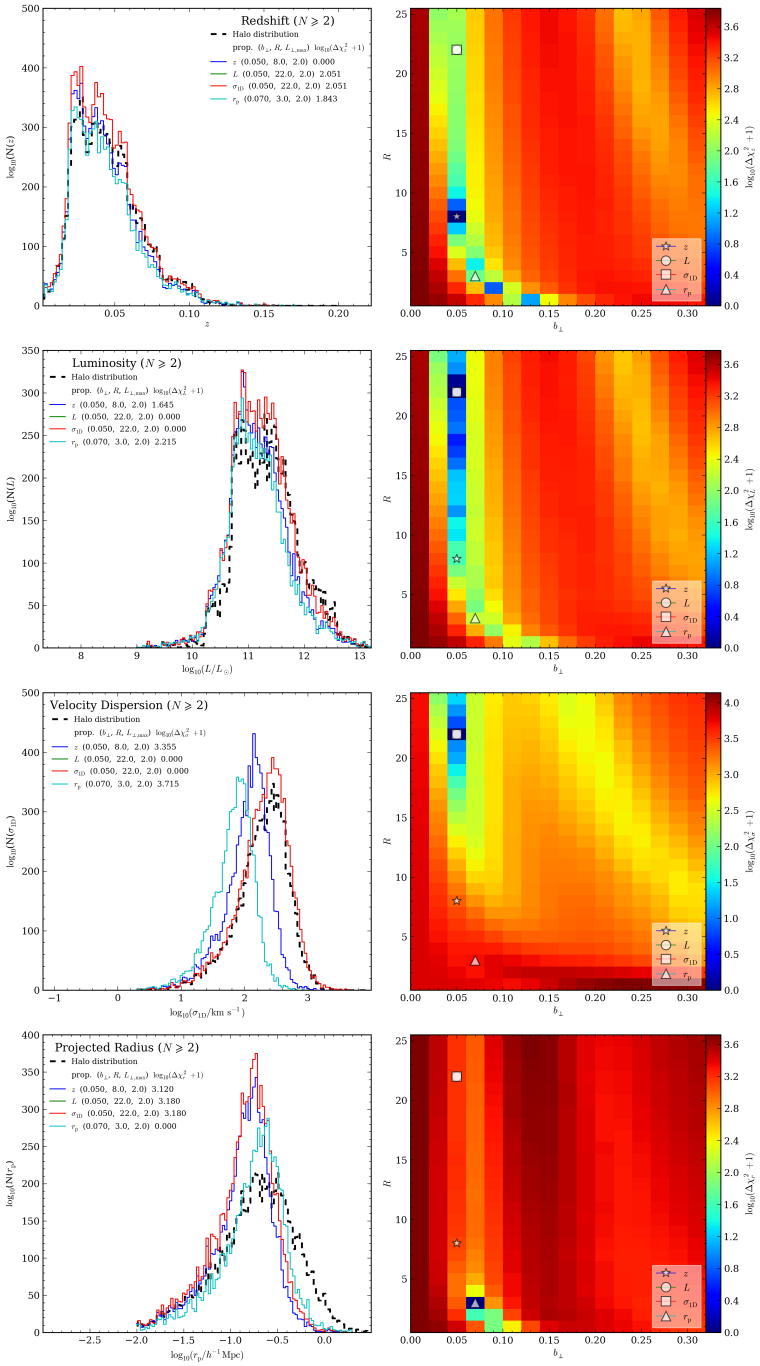

Figure 5.10: Results minimising the property the χ^2 statistic using each of the individual halo and group properties. Here we have constructed the property distributions using 100 bins. The symbols, lines and colour bar have the same meanings as in Figure 5.4.Note that the square and the circle symbols are lying on top of one another and the red line and the green line are lying on top of one another.

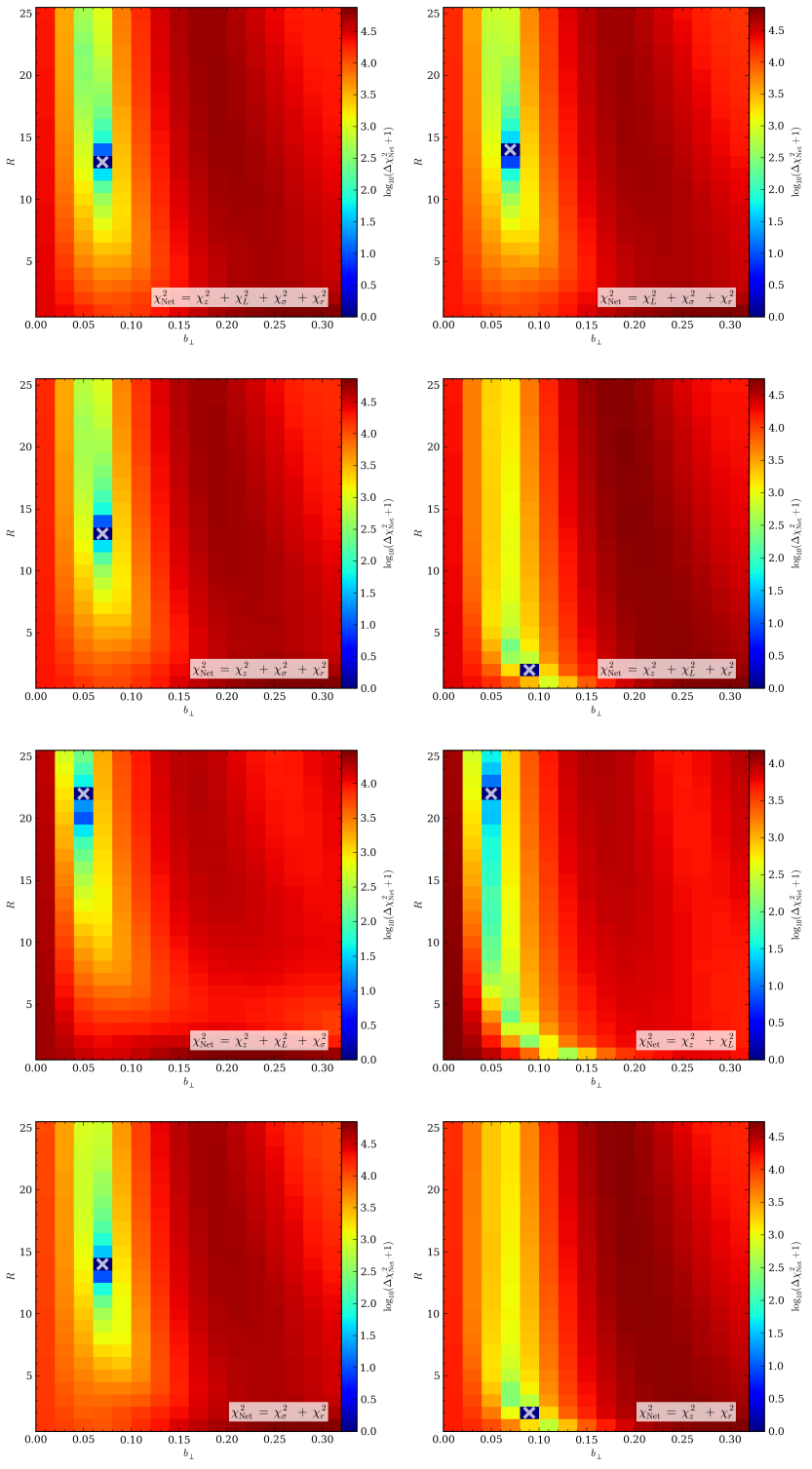

Figure 5.11: Pixel maps showing the net value for χ^2 obtained when summing the χ^2 values for different combinations of properties. Here we have constructed the property distributions using 20 bins. The symbols and colour bars have the same meaning as in Figure 5.7.

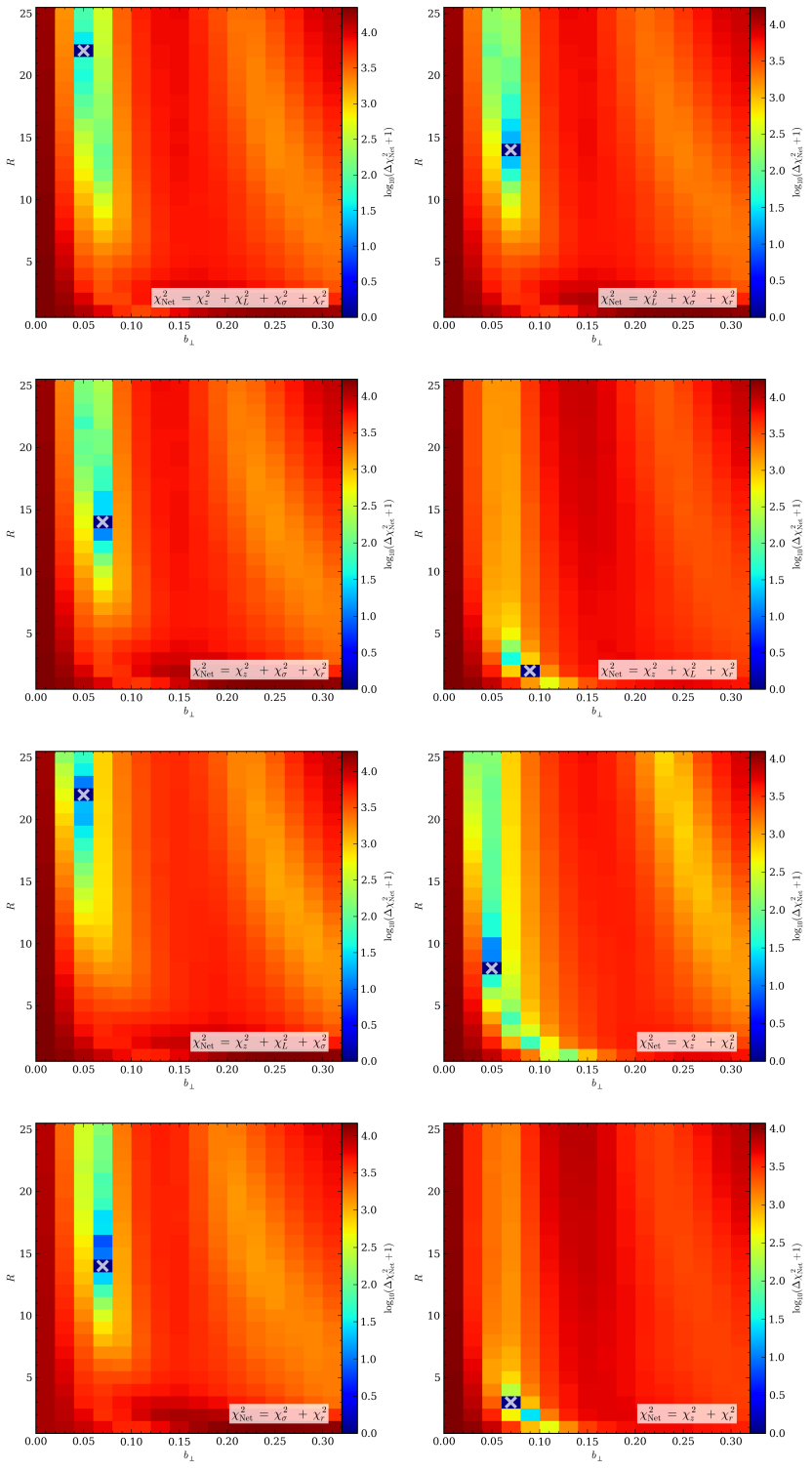

Figure 5.12: Pixel maps showing the net value for χ^2 obtained when summing the χ^2 values for different combinations of properties. Here we have constructed the property distributions using 100 bins. The symbols and colour bars have the same meaning as in Figure 5.7.

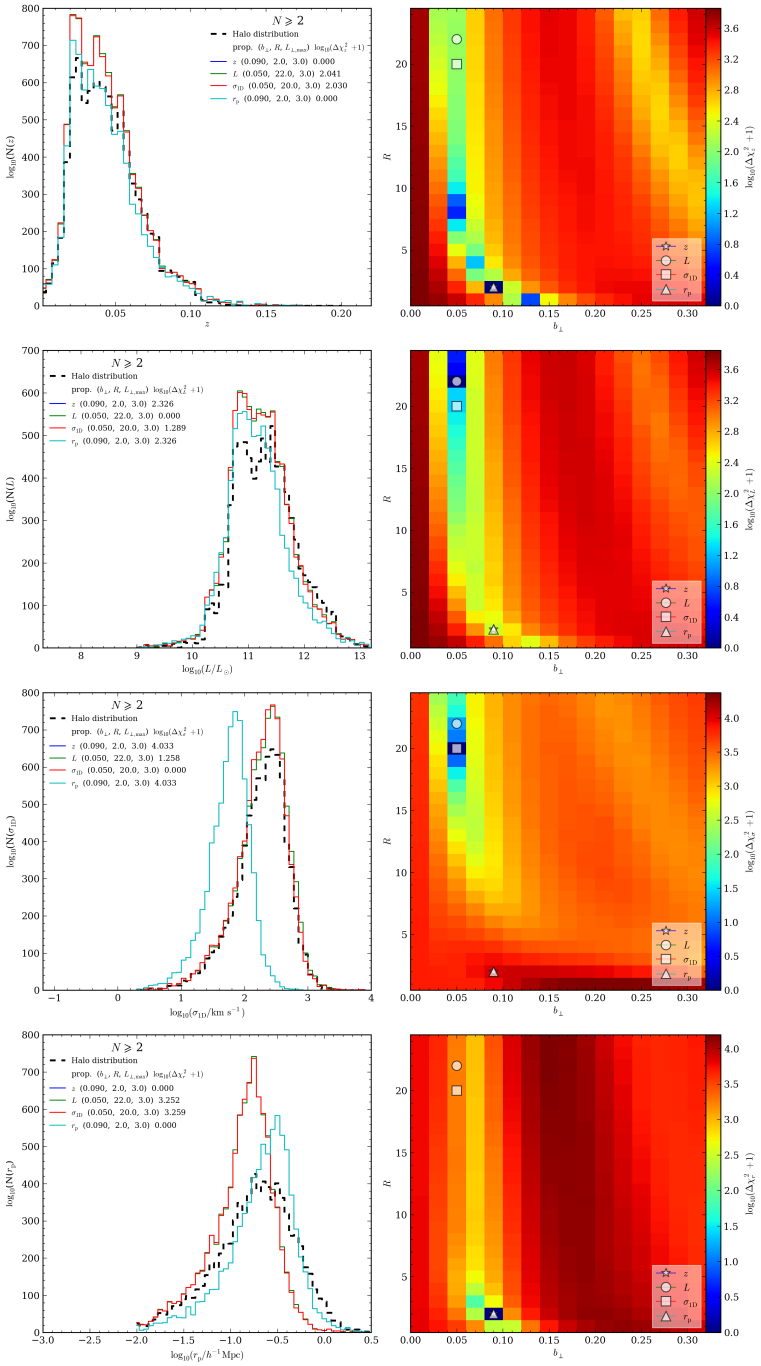

Figure 5.13: Results minimising the property the χ^2 statistic using each of the individual halo and group properties. Here we have chosen to fix $L_{\perp,\max} = 3\,h^{-1}\mathrm{Mpc}$. The symbols, lines and colour bar have the same meanings as in Figure 5.4. Note that the star and the triangle symbols are lying on top of one another and that, therefore, the blue and the cyan lines are also lying on top of one another.

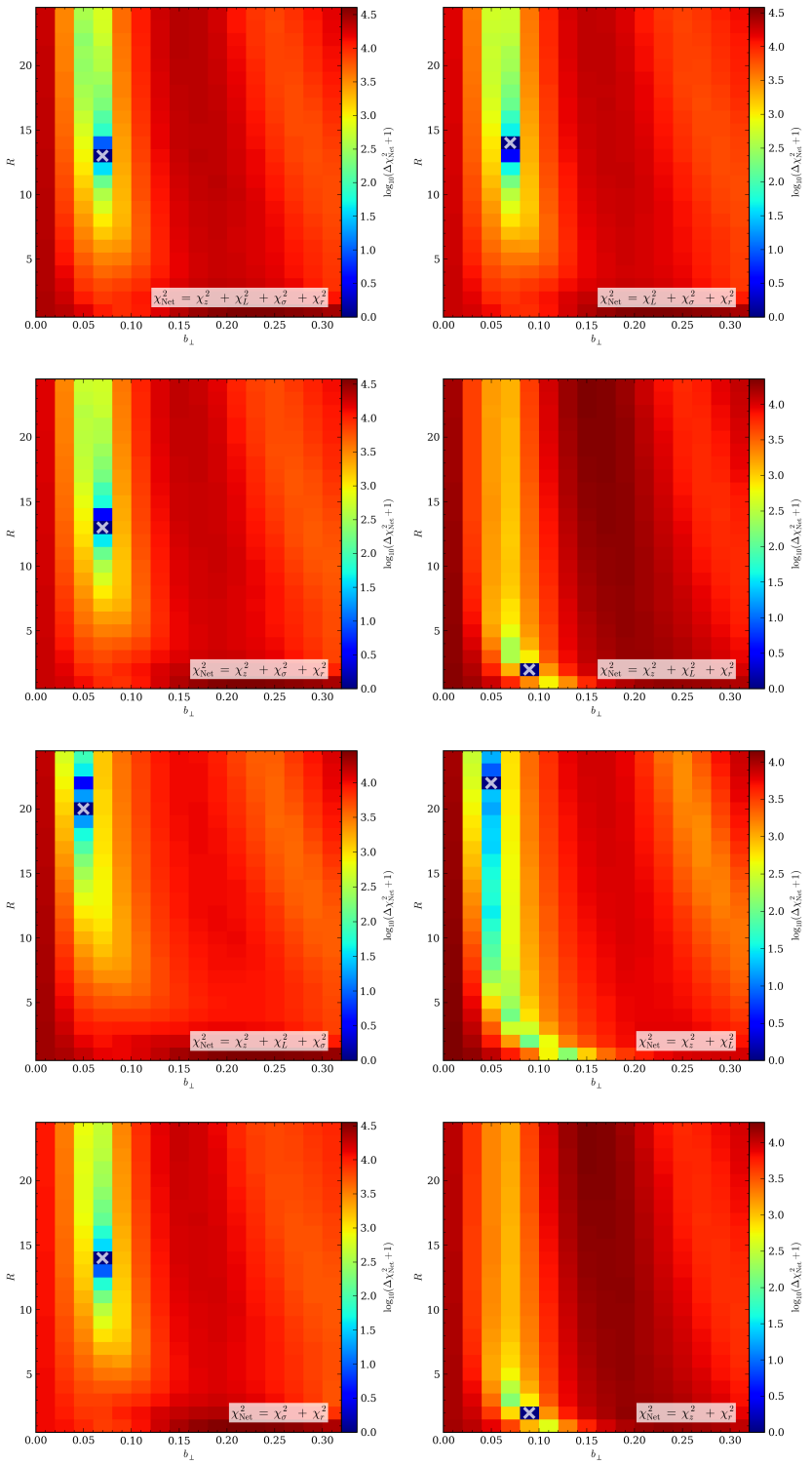

Figure 5.14: Pixel maps showing the net value for χ^2 obtained when summing the χ^2 values for different combinations of properties. Here we have chosen to fix $L_{\perp,\max} = 3\,h^{-1}$Mpc. The symbols and colour bars have the same meaning as in Figure 5.7.

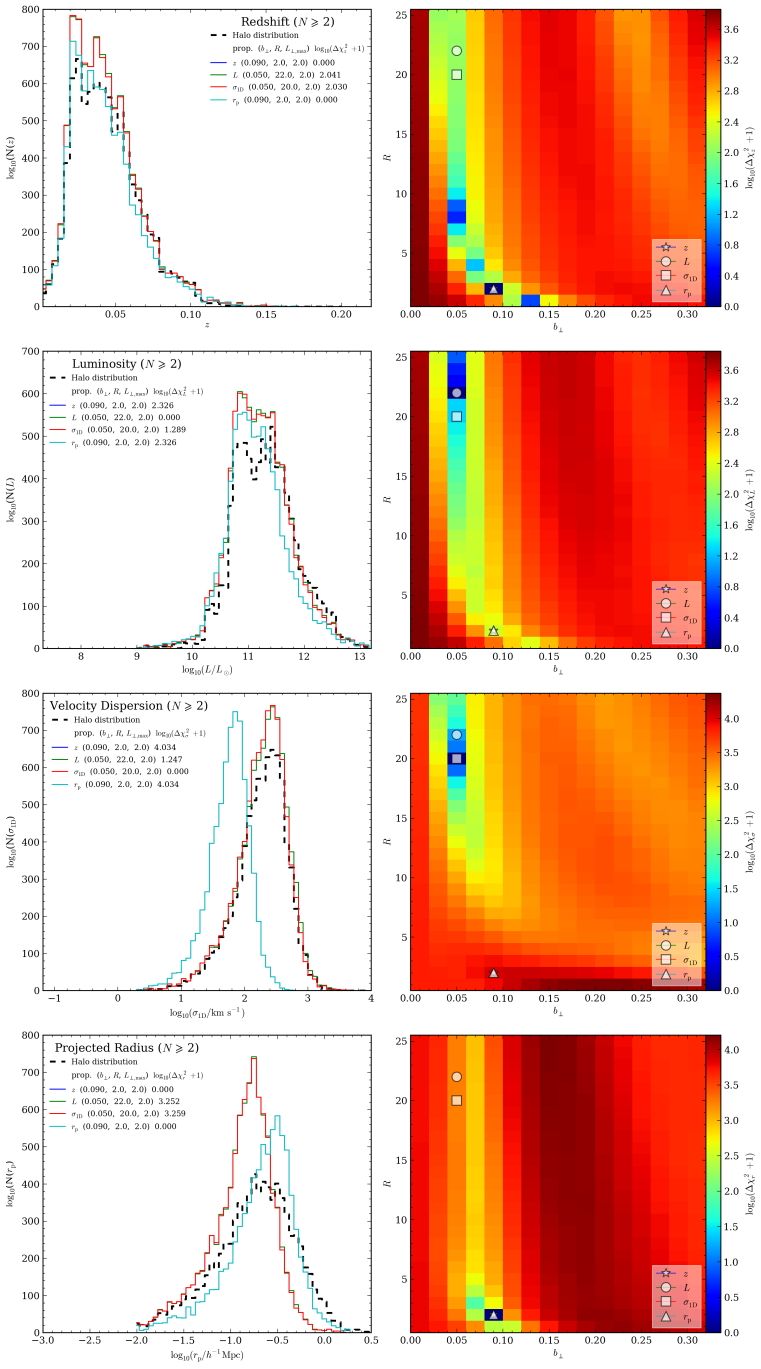

Figure 5.15: Results minimising the property the χ^2 statistic using each of the individual halo and group properties. Here we have chosen to calculate χ^2 by only using bins in the property distribution that contain at least 10 halos (corresponding to an error of approximately 30 per cent). The symbols, lines and colour bar have the same meanings as in Figure 5.4. Note that the star and the triangle symbols are lying on top of one another and that, therefore, the cyan and the blue lines are also lying on top of one another.

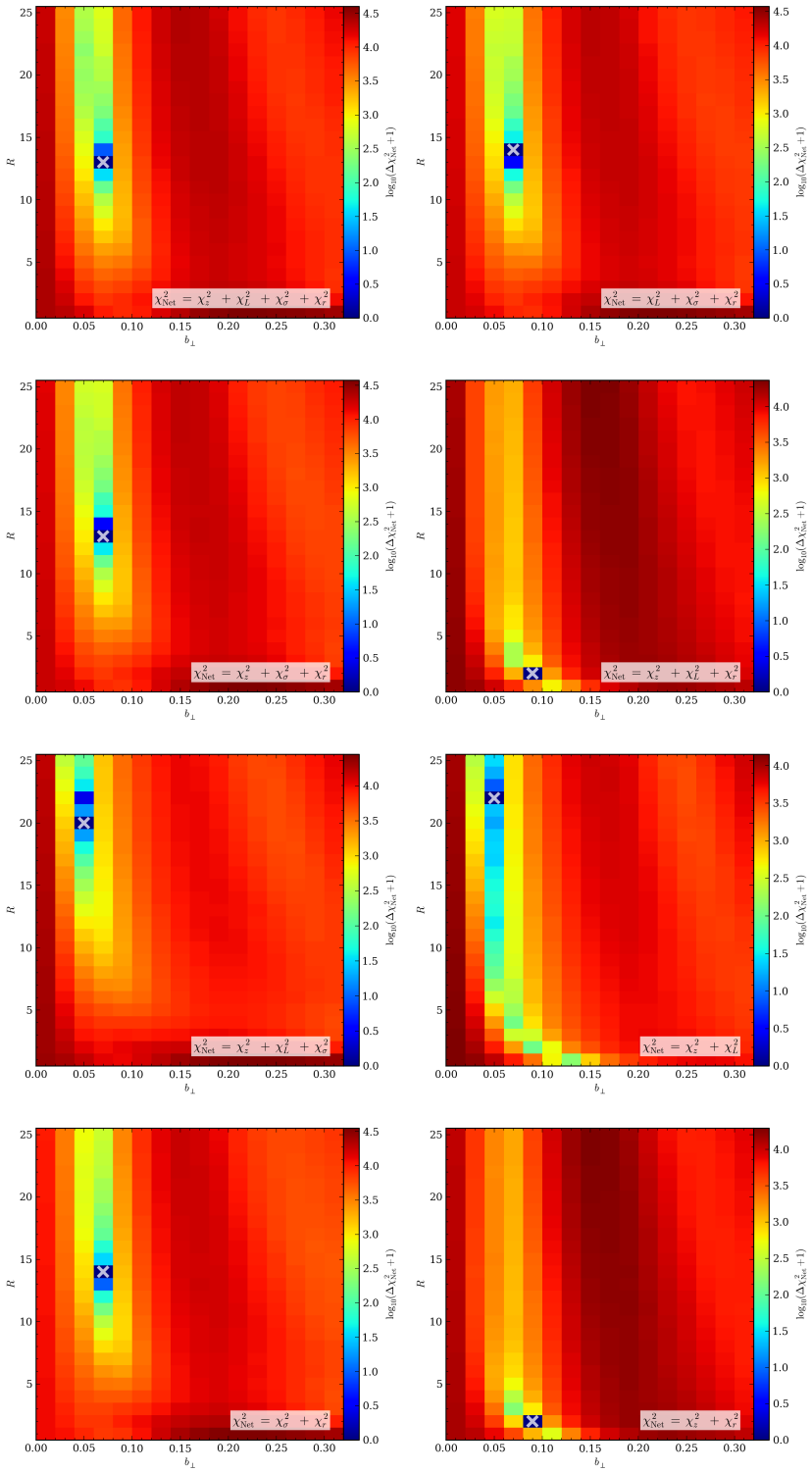

Figure 5.16: Pixel maps showing the net value for χ^2 obtained when summing the χ^2 values for different combinations of properties. Here we have chosen to calculate χ^2 by only using bins in the property distribution that contain at least 10 halos (corresponding to an error of approximately 30 per cent). The symbols and colour bars have the same meaning as in Figure 5.7.

the bin error, as we are only restricting the redshift range used for the comparison (not the range in the other properties).

We therefore repeat the comparison of the property distributions, restricting the groups and halos we use to those lying in the redshift range $0.03 \leqslant z \leqslant 0.07$. We show in Figure 5.17 and Figure 5.18 the results of this comparison. Restricting the redshift range shifts the values of the optimum linking lengths recovered when considering each property separately. The values are now distributed along the degeneracy at small values of $b_{\perp}$. When calculating the net value for χ^2, we find that restricting the redshift range has only a modest affect on the values of the recovered linking lengths. This is as we would expect since the tolerance we are applying in our χ^2 calculation is already weighting towards recovery of the peak of the distribution. Using just the velocity dispersion and the projected sizes, the values $(b_{\perp}, R) = (0.07, 14)$ are recovered as before. If either redshift or luminosity are added into the calculation, then the values shift slightly to $(b_{\perp}, R) = (0.07, 13)$.

5.6.7 Effect of multiplicity

We naturally expect that the limiting richness of the groups that we select in order to carry out the comparison will affect the values of the optimum linking lengths. For instance, if we compare the distributions for all groups and halos hosting at least 10 members that are visible above the flux limit, we will recover larger linking lengths than if we only consider systems that have a minimum of 2 or 3 members that are visible.

To examine the extent to which our optimum linking lengths change as a result of group richness we compare the distributions again, this time using limiting richnesses of $N_{\rm obs} \geqslant 3$, $N_{\rm obs} \geqslant 5$ and $N_{\rm obs} \geqslant 10$ members.

In Figure 5.19 and Figure 5.20 we show the calibration results obtained when adopting an uncorrected richness limit of 3, i.e. only considering groups/halos that host at least 3 members visible above the flux limit. In most cases, the optimum linking lengths returned for each property separately are a better match to the corresponding halo distribution. If we calculate the net χ^2 values, we see that the optimum linking lengths returned are larger. Most have $b_{\perp} = 0.07$ but the values of R vary between 16 and 25. We plot the property distributions for the optimum linking lengths in Figure 5.21.

In Figure 5.22 and Figure 5.23 we show the calibration results adopting a minimum richness of 5. Considering first the individual properties, the minimum χ^2 values are much smaller and the recovered group catalogues at the corresponding linking lengths are

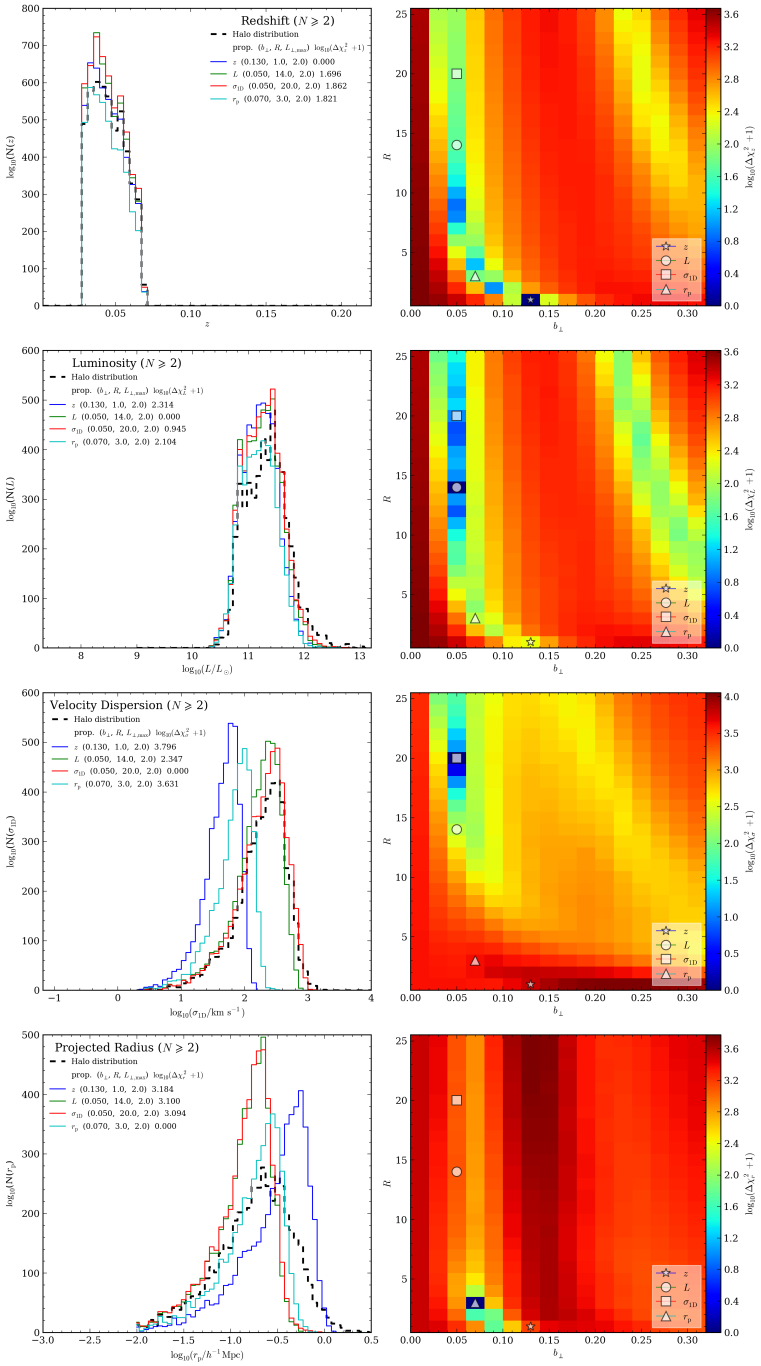

Figure 5.17: Results minimising the property the χ^2 statistic using each of the individual halo and group properties. Here we have chosen to calculate χ^2 by only using the properties of halos and groups with redshifts within the range $0.03 \leqslant z \leqslant 0.07$. The symbols, lines and colour bar have the same meanings as in Figure 5.4.

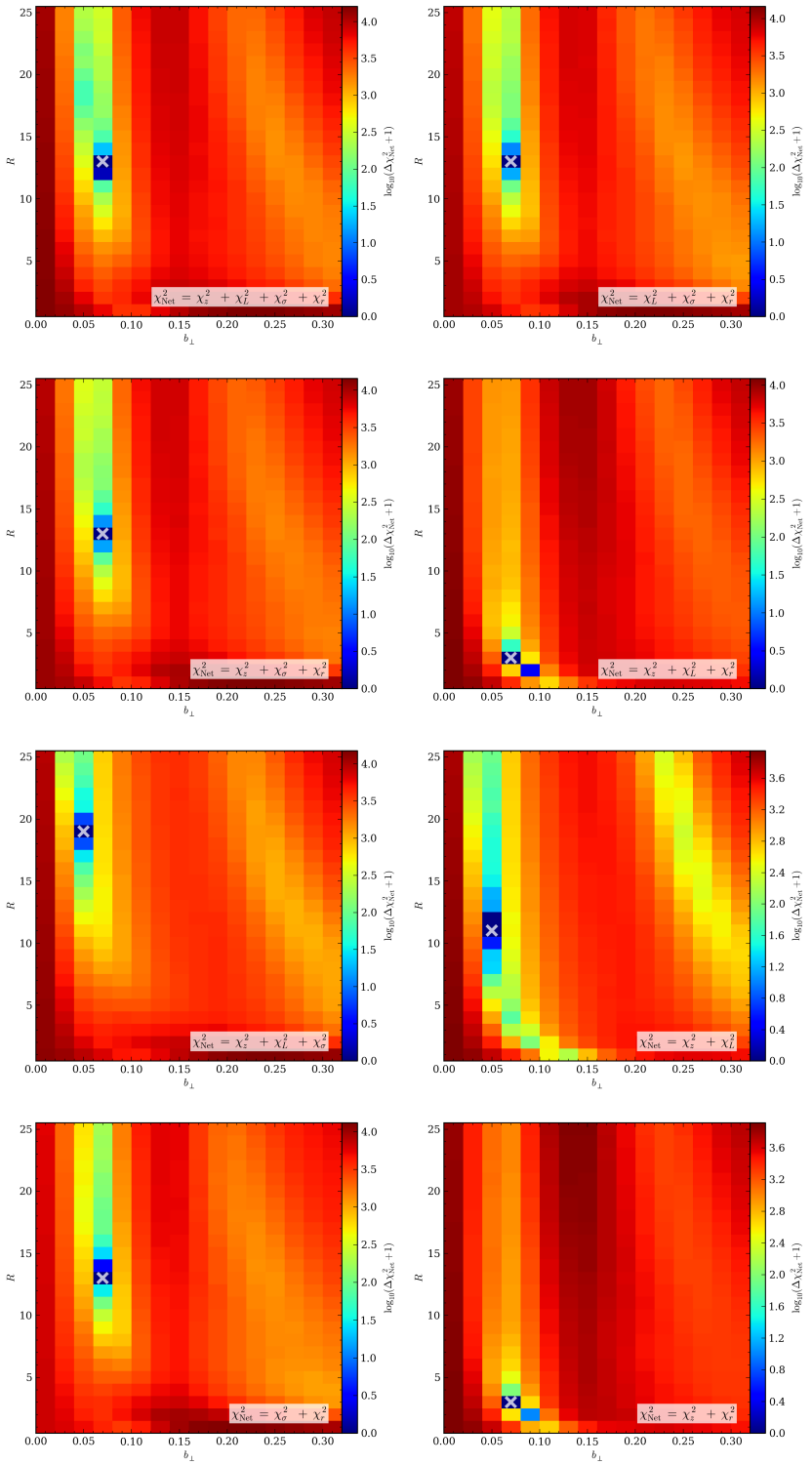

Figure 5.18: Pixel maps showing the net value for χ^2 obtained when summing the χ^2 values for different combinations of properties. Here we have chosen to calculate χ^2 by only using the properties of halos and groups with redshifts within the range $0.03 \leqslant z \leqslant 0.07$. The symbols and colour bars have the same meaning as in Figure 5.7.

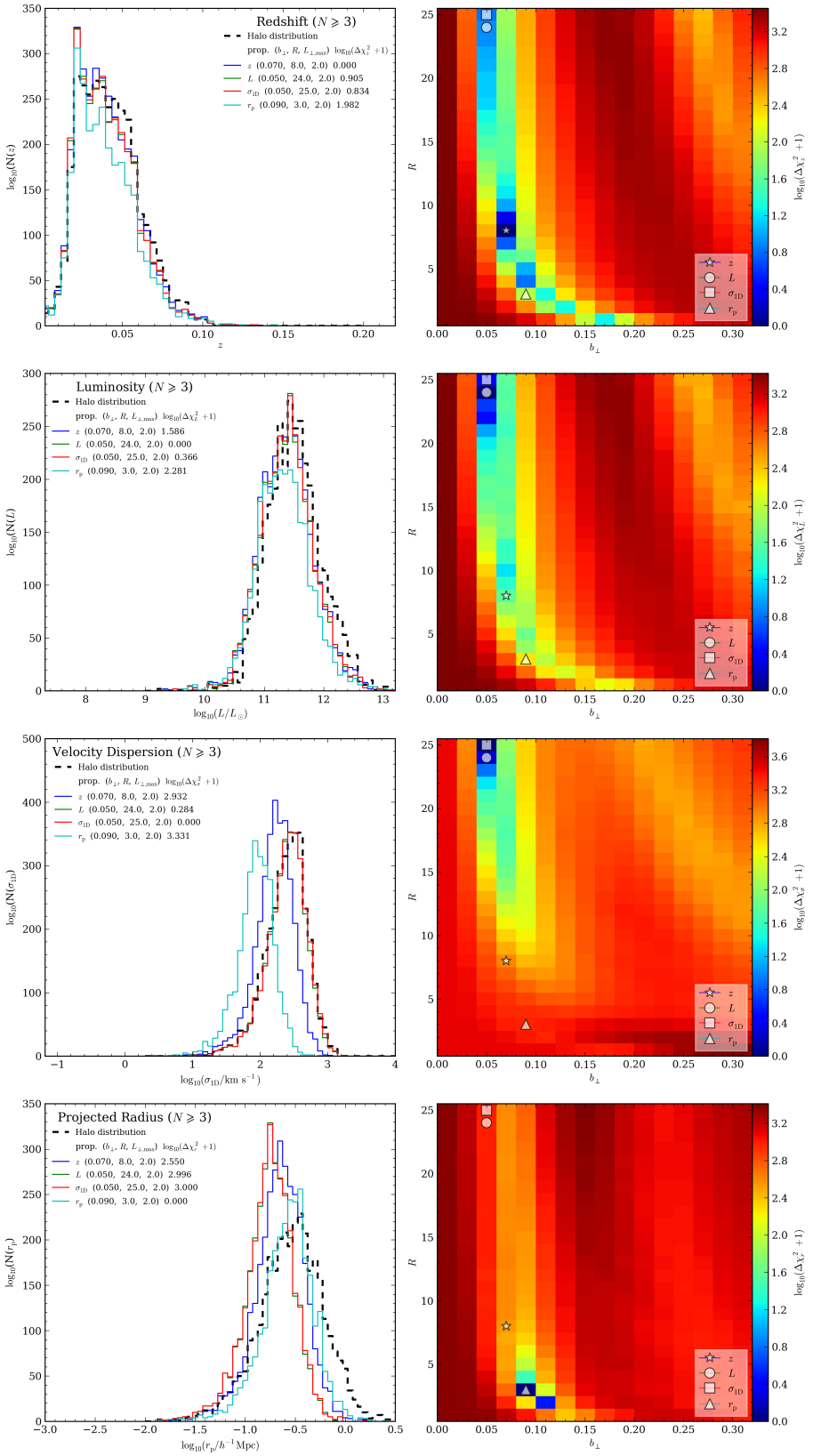

Figure 5.19: Results minimising the property the χ^2 statistic using each of the individual halo and group properties. Here we consider only halos and groups that host at least 3 galaxies that are visible above the flux limit. The symbols, lines and colour bar have the same meanings as in Figure 5.4.

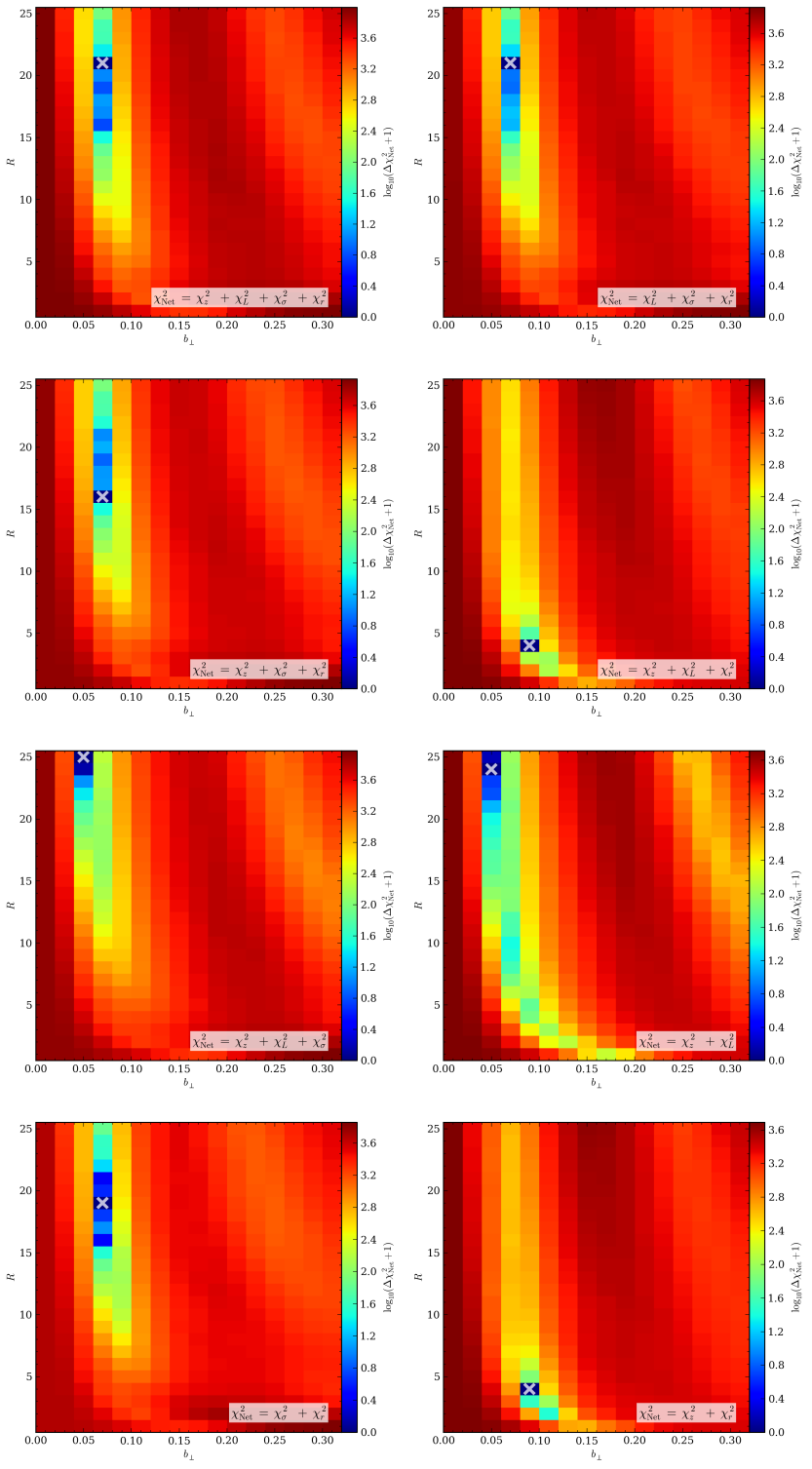

Figure 5.20: Pixel maps showing the net value for χ^2 obtained when summing the χ^2 values for different combinations of properties. Here we consider only halos and groups that host at least 3 galaxies that are visible above the flux limit. The symbols and colour bars have the same meaning as in Figure 5.7.

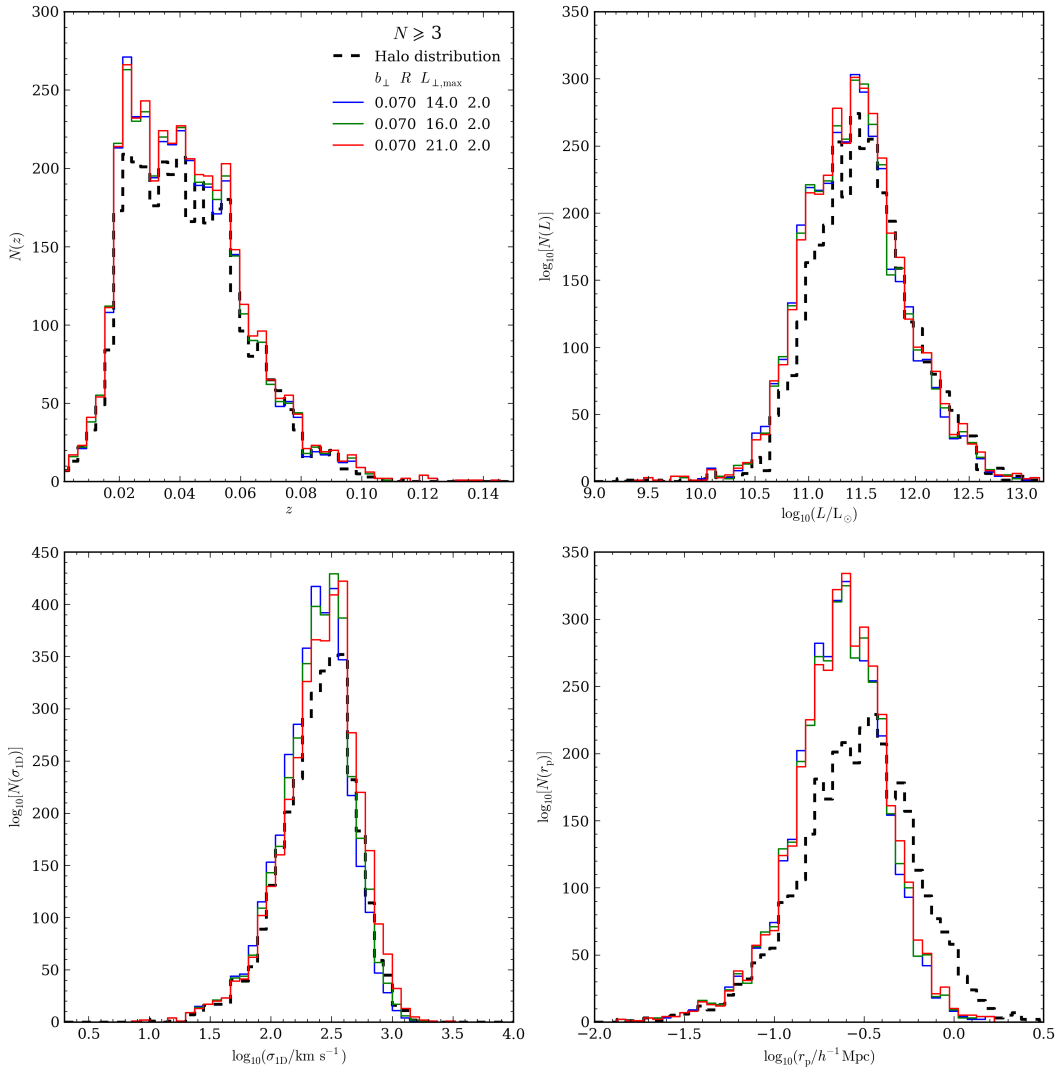

Figure 5.21: Property distributions for galaxy groups recovered using different linking length parameter combinations, as shown by the various coloured lines. Here we consider only halos and groups that host at least 3 galaxies that are visible above the flux limit. The four panels correspond to the same group properties shown in Figure 5.8.

an excellent fit to the corresponding halo distributions. Interestingly, there is a change around in the pairs of properties that are in agreement. With $N_{\rm obs} \geqslant 5$, the optimum linking lengths recovered using the redshift alone, are now in agreement with those from the velocity dispersion, whilst the linking lengths returned using the luminosity are in better agreement with those from the projected sizes. If we now examine the net χ^2 values obtained by combining the different properties, we find that there is much better agreement and in all of the cases where both the velocity dispersion and projected sizes are considered, with the possible addition of both or one of the redshift and the luminosity, the optimum linking lengths are always $(b_{\perp}, R) = (0.07, 21)$. In Figure 5.24 we plot the property distributions for groups recovered using $(b_{\perp}, R) = (0.07, 21)$ as well as for groups recovered using $(b_{\perp}, R) = (0.07, 14)$. The distributions for the majority of group properties obtained using these two sets of linking lengths both appear to be consistent with the corresponding distributions for the halos, as well as with each other.

In Figure 5.25 and Figure 5.26 we show the calibration results obtained when adopting a minimum richness of 10. Note that due to the reduced number of systems with at least 10 galaxies visible above the flux limit, we have increased the tolerance on our χ^2 calculation to 30 per cent. If we consider the linking lengths from the separate properties, we see that the χ^2 values are again lower and the values of the linking lengths for the various properties are beginning to approach one another in the parameter space. We see that this is the first instance in which the linking lengths recovered by calibrating using the velocity dispersions provide a reasonable match to the distribution of projected sizes and vice versa. Considering now the net χ^2 values, we see that the values of the optimum linking lengths are understandably larger, with $b_{\perp} = 0.09$ and $R \sim 22 - 24$. In Figure 5.27 we plot the property distributions for the predicted optimum linking lengths from Figure 5.26 and compare them to the distributions if we adopt $(b_{\perp}, R) = (0.07, 14)$. Understandably, this latter linking length combination under-predicts the number of groups for all four properties, since it was predicted as the optimum combination when using a smaller richness limit. The two values predicted for a richness limit of 10, provide a good match to all of the property distributions.

Increasing the minimum multiplicity clearly increases the size of the optimum linking lengths recovered. Although the optimum linking lengths for the different properties begin to converge for high multiplicities, the preferred linking lengths for the survey as a whole will depend upon the multiplicity of the groups one wishes to prioritise, i.e. the recovery of large or small groups. From this we conclude that it is extremely difficult to obtain a set

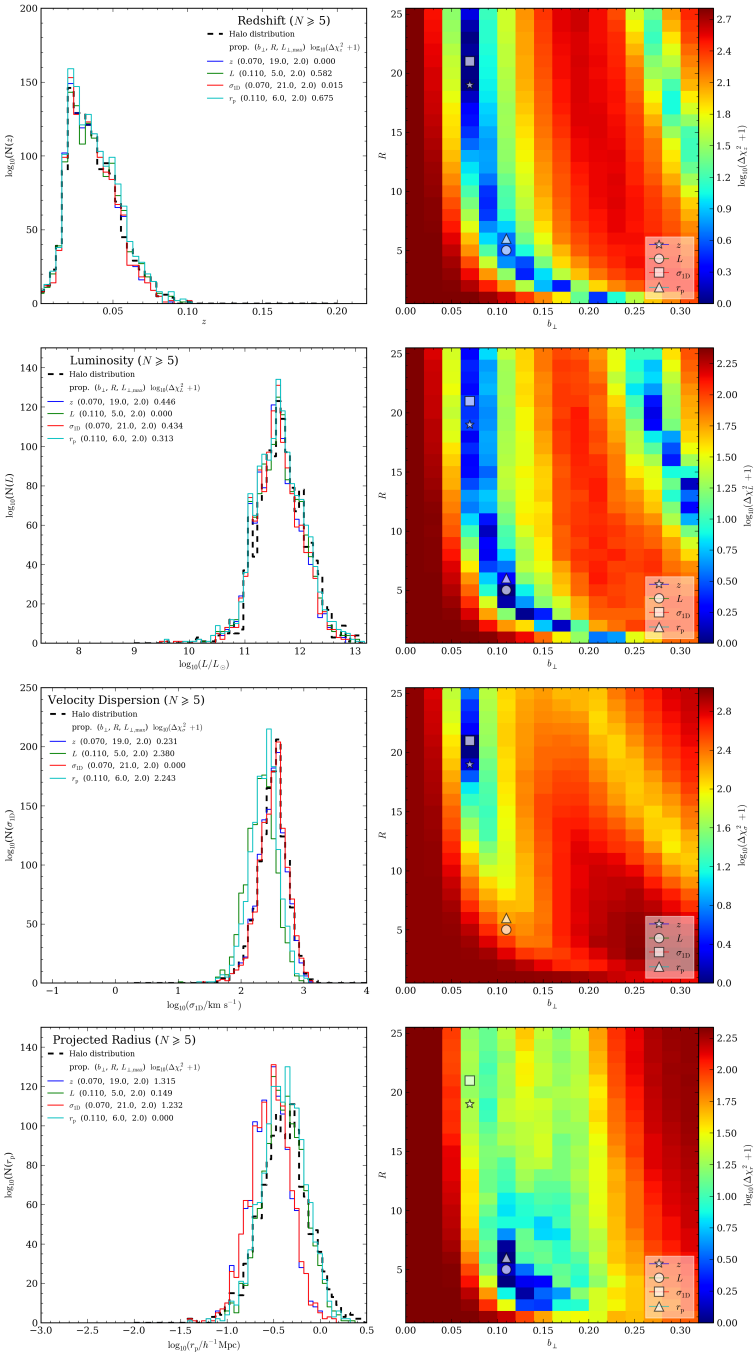

Figure 5.22: Results minimising the property the χ^2 statistic using each of the individual halo and group properties. Here we consider only halos and groups that host at least 5 galaxies that are visible above the flux limit. The symbols, lines and colour bar have the same meanings as in Figure 5.4.

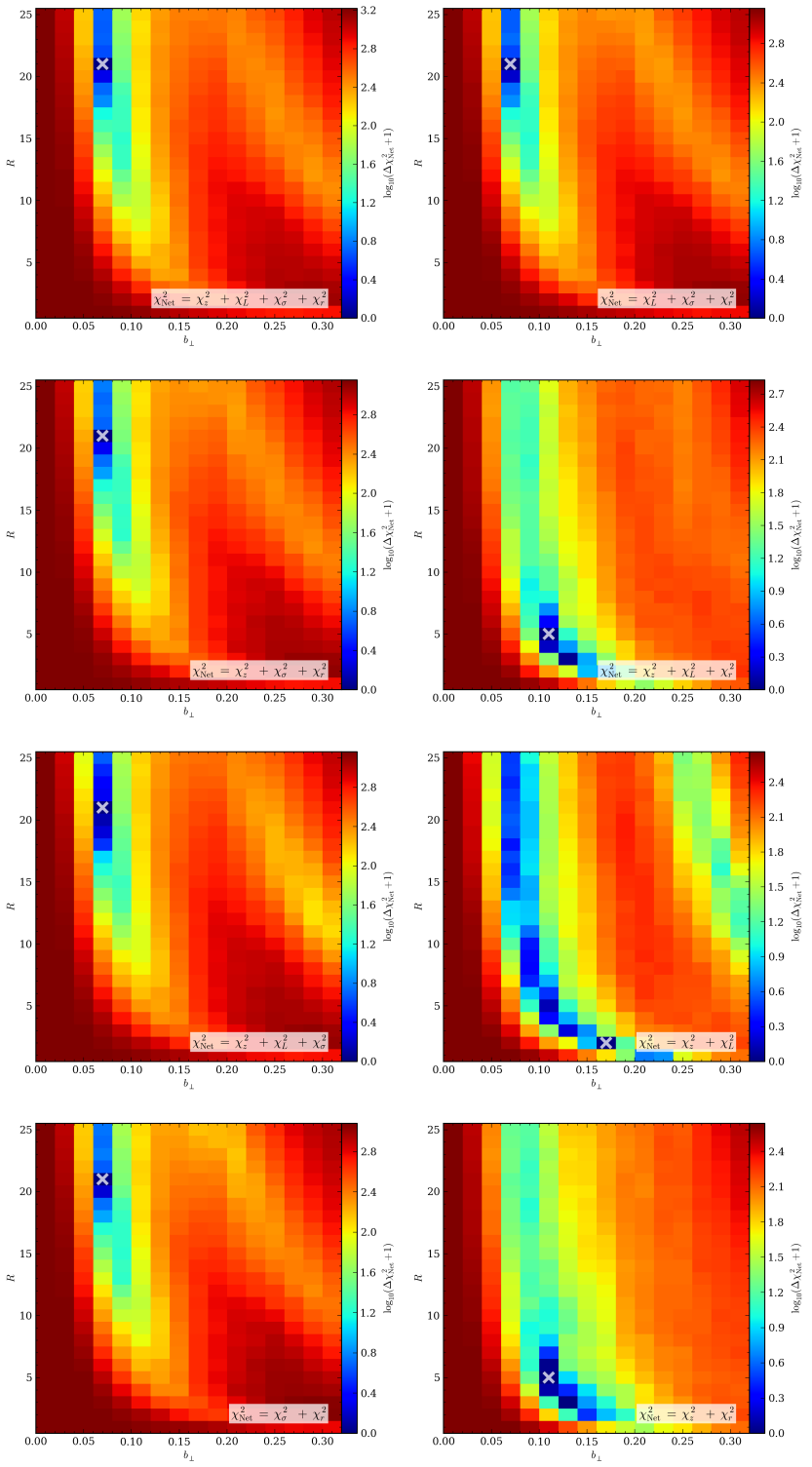

Figure 5.23: Pixel maps showing the net value for χ^2 obtained when summing the χ^2 values for different combinations of properties. Here we consider only halos and groups that host at least 5 galaxies that are visible above the flux limit. The symbols and colour bars have the same meaning as in Figure 5.7.

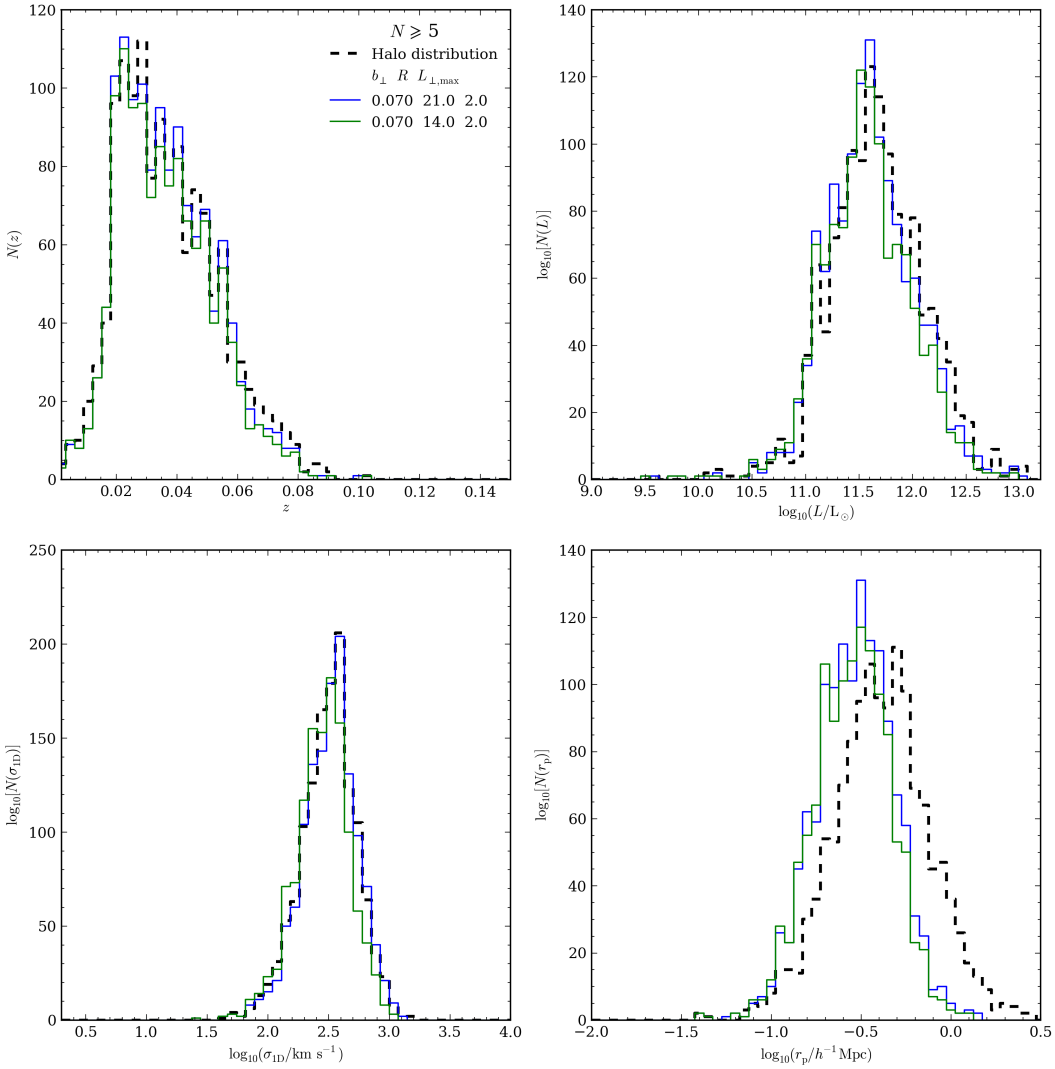

Figure 5.24: Property distributions for galaxy groups recovered using different linking length parameter combinations, as shown by the various coloured lines. Here we consider only halos and groups that host at least 5 galaxies that are visible above the flux limit. The four panels correspond to the same group properties shown in Figure 5.8.

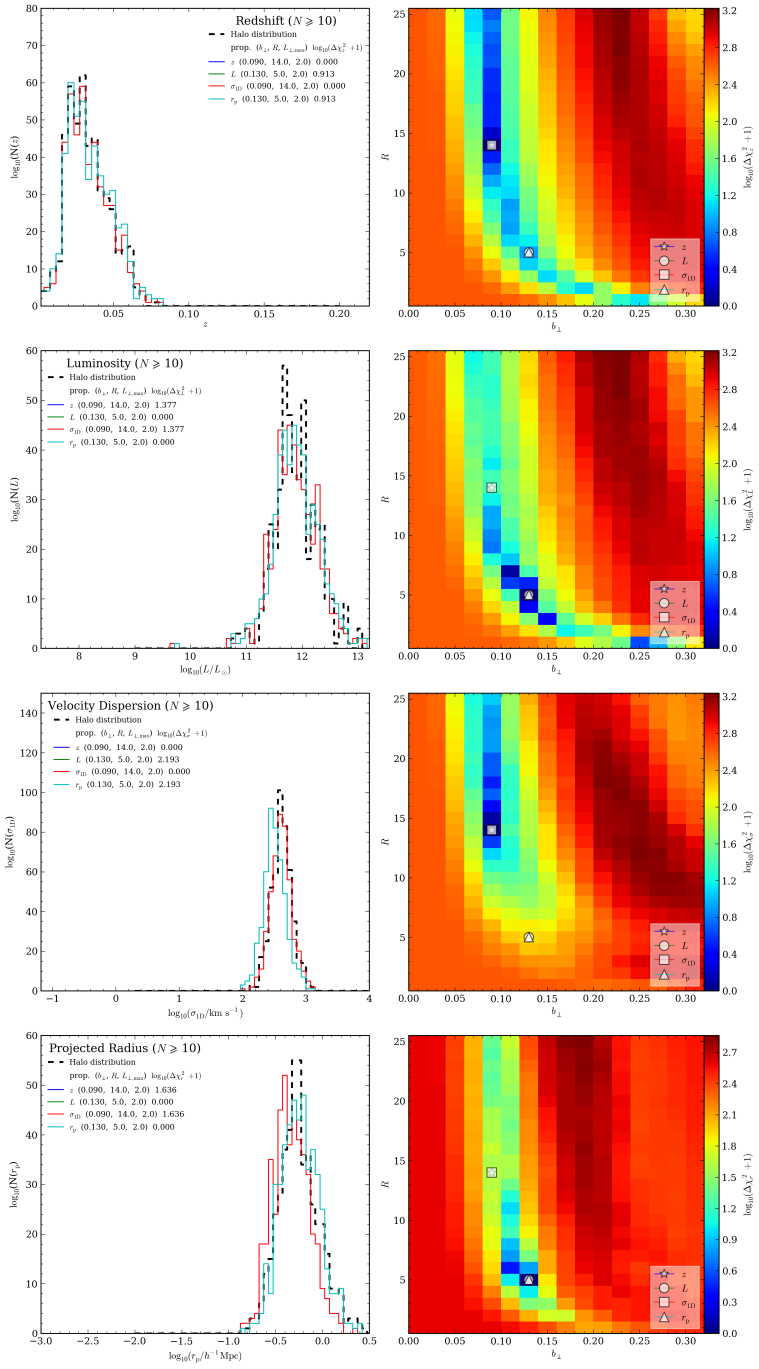

Figure 5.25: Results minimising the property the χ^2 statistic using each of the individual halo and group properties. Here we consider only halos and groups that host at least 10 galaxies that are visible above the flux limit. The symbols, lines and colour bar have the same meanings as in Figure 5.4. Note that the square and the star are lying on top of one another, the circle and the triangle are lying on top of one another, the blue line and the red line are lying on top of one another and the green line and the cyan line are lying on top of one another.

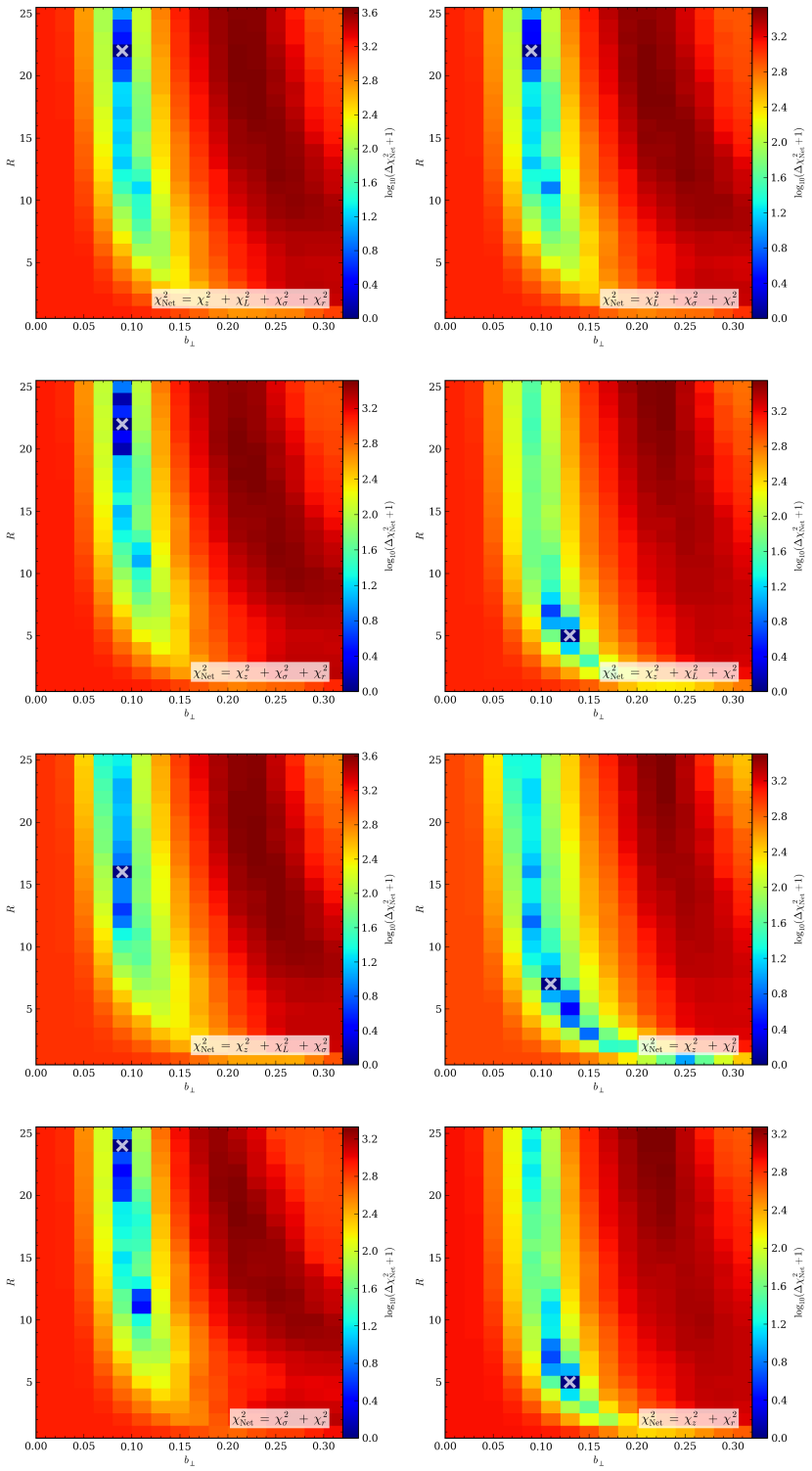

Figure 5.26: Pixel maps showing the net value for χ^2 obtained when summing the χ^2 values for different combinations of properties. Here we consider only halos and groups that host at least 10 galaxies that are visible above the flux limit. The symbols and colour bars have the same meaning as in Figure 5.7.

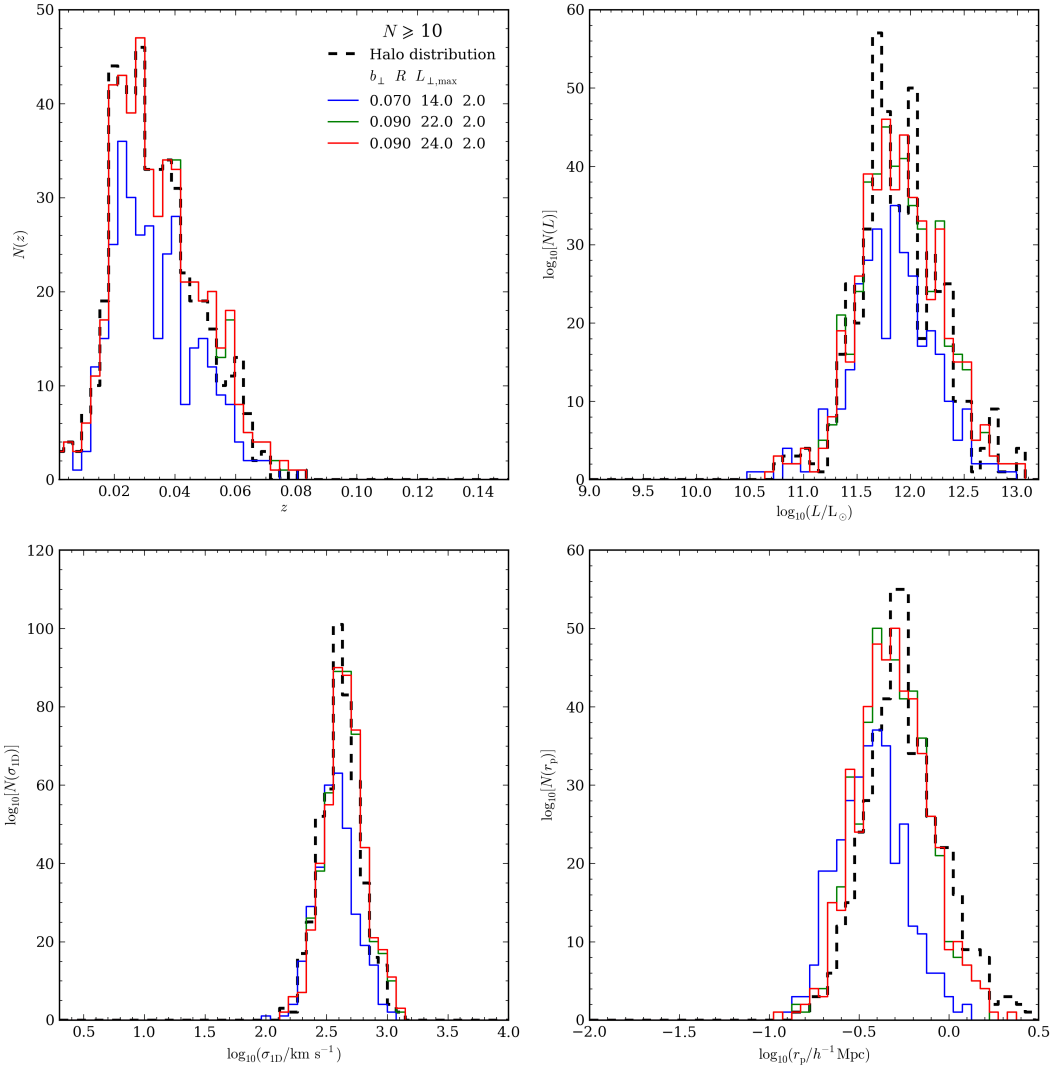

Figure 5.27: Property distributions for galaxy groups recovered using different linking length parameter combinations, as shown by the various coloured lines. Here we consider only halos and groups that host at least 10 galaxies that are visible above the flux limit. The four panels correspond to the same group properties shown in Figure 5.8.

of linking length parameters that simultaneously recover all of the property distributions for low richness systems. For high richness systems it is possible to begin to converge on a single set of linking lengths, though high richness systems will contribute a smaller fraction of the total groups catalogue. One option would be to construct multiple groups catalogues with each catalogue targeting the optimum recovery of groups in different bins of group richness.

5.6.8 Effect of luminosity limit

If we adopt a different luminosity limit, corresponding to $M^\star + 2$, then the optimum linking lengths recovered do not change.

Throughout this analysis we have adopted to use a luminosity limit corresponding to $M^\star + 1$. To do this, we corrected the observed luminosities by integrating over a Schechter fit to the 6dFGS luminosity function (see Section 5.4.2) and multiplying the observed group luminosities by a correction factor, f. Larger values of f mean that we are missing a larger fraction of the light from a group. So far we have not placed any restriction on the value of f. However, it is not ideal to be considering halos and groups in our calibration where we are only seeing less than half of the light from that group, i.e. where $f \geqslant 2$. We therefore check whether, if we exclude halos and groups for which $f \geqslant 2$, our values for the optimum linking lengths change. Figure 5.28 and Figure 5.29 show the results of our calibration if we retain the luminosity limit $M^\star + 1$, but only consider halos/groups with $f < 2$. As we can see from Figure 5.29, restricting the halos/groups to ones for which we can see over 50 per cent of the light, has no appreciable affect upon the values that we recover for the optimum linking lengths.

5.7 Conclusions & outlook

In this chapter we have calibrated a group-finder for use with the 6dFGS. To do this we have built a 6dFGS mock catalogue using the `GALFORM` model and rescaled the predicted galaxy magnitudes to provide an excellent match to the 6dFGS luminosity function. We combined the `GALFORM` output with FOF halos, drawn from the *Millennium Simulation* (rather than the post-processed halos described in Chapter 2), to match the percolation algorithm used to find galaxy groups.

In this work we have chosen to calibrate a FOF galaxy group-finder by searching for the linking length parameters that provide the group catalogue that best matches the

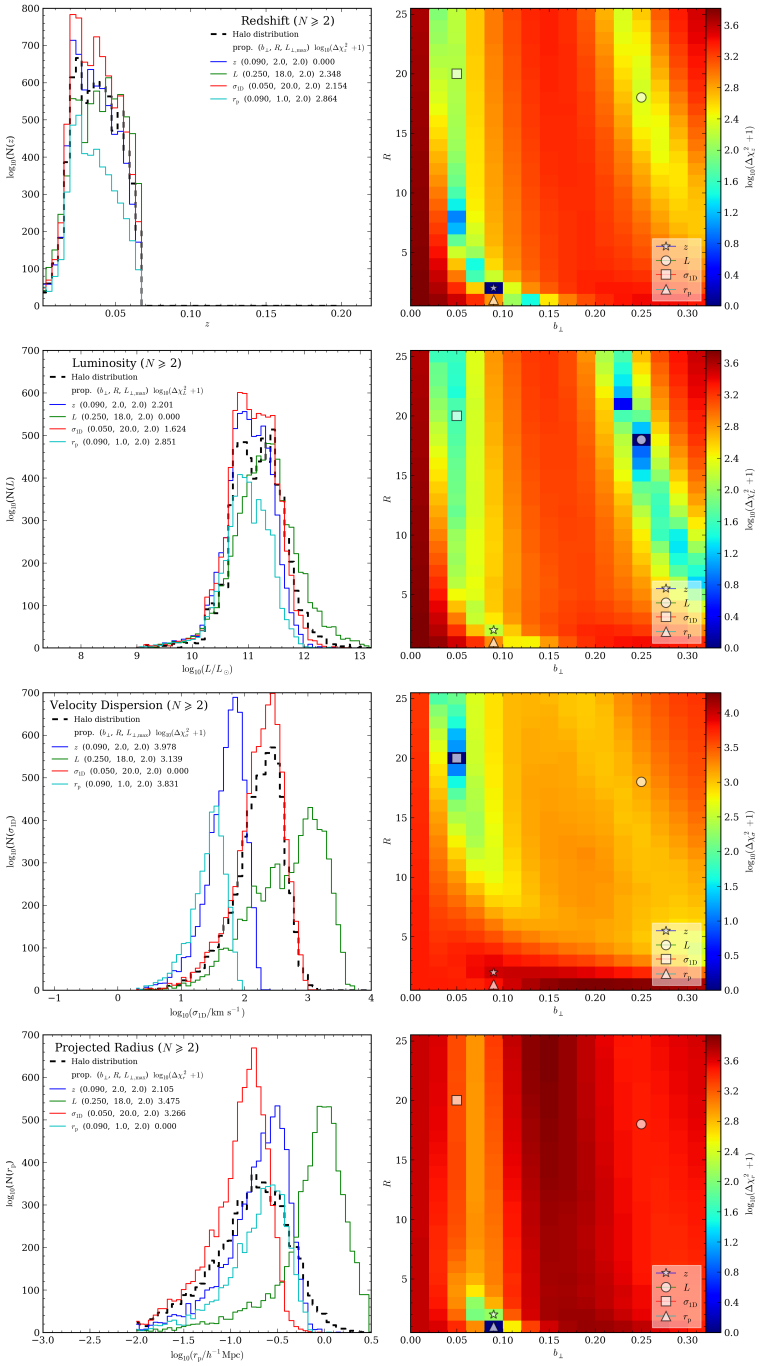

Figure 5.28: Results minimising the property the χ^2 statistic using each of the individual halo and group properties. Here we consider only halos and groups for which greater than 50 per cent of the light from the system is visible prior to applying the correction factor. The symbols, lines and colour bar have the same meanings as in Figure 5.4.

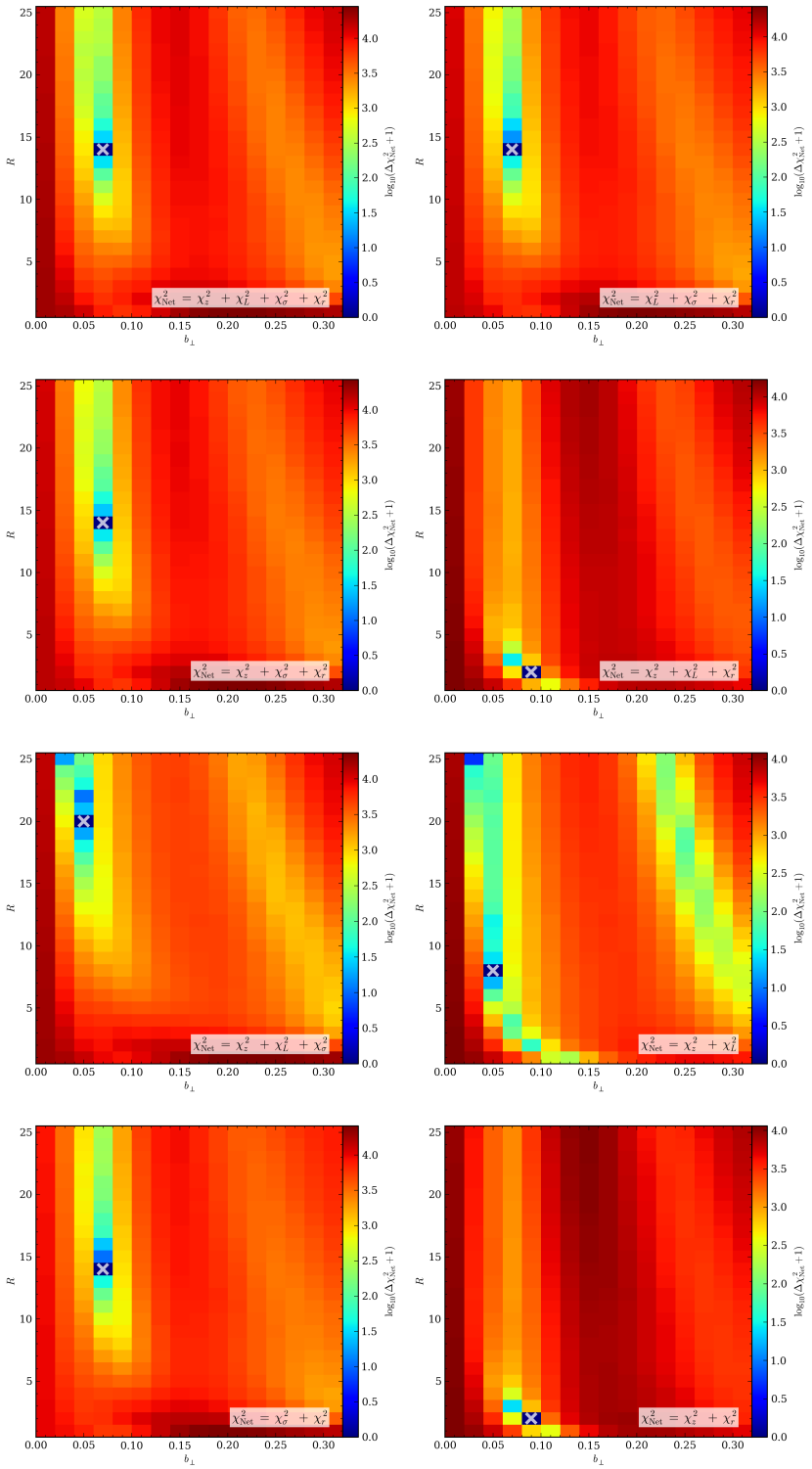

Figure 5.29: Pixel maps showing the net value for χ^2 obtained when summing the χ^2 values for different combinations of properties. Here we consider only halos and groups for which greater than 50 per cent of the light from the system is visible prior to applying the correction factor. The symbols and colour bars have the same meaning as in Figure 5.7.

statistics of the population of halos. We use the χ^2 statistic to quantify how well the properties of the recovered groups match the true distribution of halo properties.

The group properties that we consider are: the redshift distribution of the recovered groups, the luminosity of the groups (above a fixed absolute magnitude limit), the line-of-sight velocity dispersion of the groups and the distribution of the projected radii of the groups. The advantage of using these properties is that they are, on the whole, easily measurable and provide constraints on the location, geometry and membership of groups.

Our calibration has shown that it is extremely difficult to determine the optimal linking lengths by minimising the χ^2 for each property individually. Clear degeneracies exist between $b_\perp$ and R in this case. We have shown, however, that by combining the χ^2 statistics for various properties, we can identify a set of linking lengths with which we can construct a catalogue of groups that provides a reasonable match to the distribution of halo properties.

By combining the χ^2 values for various properties, we find that the linking lengths $(b_\perp, R) = (0.07, 14)$ appear to provide a reasonable match to the majority of the properties. We have considered several factors that could affect our recovery of the optimal linking lengths. In most of the cases considered, the change in the value of the recovered linking lengths is minimal. Of the factors considered, the largest change in the linking lengths comes from changing the minimum multiplicity of the groups one wishes to recover. Increasing the minimum multiplicity leads to larger values for the optimal linking lengths. It is possible, therefore, that one would need to use different linking lengths to construct different versions of group catalogues, with each designed to faithfully recover groups of different richness or to give the best match to a particular statistic.

Our calibration has shown that it is not possible to locate a single set of linking lengths that provides optimum recovery of all halo properties, at least for groups with low richness. The two properties that are the most difficult to recover simultaneously are the velocity dispersion of groups and the projected sizes. Just like the group richness, it is likely that one would need to construct multiple groups catalogues, each using a different set of linking lengths, designed to favourably recover each of these properties in turn.

Further calibration work includes:

- Perform a one-to-one matching to check whether our method of matching the statistical distributions is effective and returns optimum linking lengths that are in agreement with the one-to-one approach.

- See how the optimum linking lengths change if we use halos that have been post-processed to look like more realistic bound structures, e.g. by removing unbound particles, similar to the construction process discussed in Chapter 2).

- Investigate further the various factors that have been shown to affect the optimal choice of linking lengths, such as the multiplicity and the luminosity of groups. Consider other group properties that might be worth incorporating into our calibration procedure.

- Consider alternative methods for comparing the halo and group distributions, such as using a cost function to match the cumulative distributions for the properties.

Our ultimate goal is to calibrate our chosen group finding algorithm for use with the 6dFGS. Based upon this preliminary work, we see the immediate future work necessary to achieve this goal as being:

1. Apply the 6dFGS completeness mask to the 6dFGS mock catalogue to ensure that the mock is as fully realistic an emulation of the survey as possible. The addition of the completeness mask will help simulate phenomena such as fibre collisions. Fibre collisions will reduce the numbers of galaxies found in dense regions on the sky. If these dense regions correspond to actual galaxy clusters, then neglecting such an effect could lead to a biased set of linking lengths being recovered. More importantly, the completeness mask will lead to a more realistic radial selection function, that will affect the number density of galaxies as a function of redshift, which will in turn affect the scaling of the linking lengths.

2. Repeat the calibration procedure for a set of 6dFGS mock catalogues, each constructed by placing the observer at different location within the *Millennium Simulation* cube and with a different orientation. This will help assess the importance that large-scale structure in the mock catalogue has upon the recovery of the optimum linking lengths.

3. A set of 6dFGS mocks should be constructed, each using a different galaxy formation model. This is necessary to help assess how sensitive our calibration procedure is to the choice of galaxy formation model.

The construction of a 6dFGS groups catalogue will allow much future work to be considered:

- The HI predictions made by the `GALFORM` model will allow predictions of the HI content of galaxy groups in the local Universe, which will be of use to the upcoming ASKAP HI surveys, *WALLABY* and *DINGO*.

- Measurement of the galaxy-group cross-correlation function to examine the distribution of galaxies within groups and clusters in the local Universe.

- The large solid angle of the 6dFGS means that a large groups catalogue covering the extremely low-redshift Universe could place tight constraints on the HOD at the present day.

- A preliminary 6dFGS groups catalogue constructed using the 2dFGRS calibrated linking lengths, $(b_{\perp}, R) = (0.13, 11)$, has been used to examine the affect of galaxy environment on the fundamental plane of local elliptical galaxies (Magoulas et al., 2012). A properly calibrated 6dFGS groups catalogue will be useful for confirming these results.

Chapter 6

Angular clustering in mock galaxy catalogues

Following on from the mapping of the low-redshift Universe by galaxy surveys such as the 2dFGRS and the SDSS, the goal of the next generation of galaxy surveys is to probe the high redshift ($z \sim 2$) galaxy population to a similar level of precision.

An obvious reason for this is to assess how the properties of the galaxy population evolve with redshift and whether the trends observed in the local galaxy population are also present at high redshift. Such observations could also be used to place tighter constraints on galaxy formation models and our predictions of how the interplay between the various processes governing galaxy formation evolves with redshift. Although pencil beam surveys, such as the DEEP2 survey (Davis et al., 2003; Newman et al., 2012) and the Virmos VLT Deep Survey (Le Fèvre et al., 2003, 2004) have already begun to examine high redshift galaxies, the small solid angles covered by these surveys means that they have opened only a small window into the high redshift Universe. Many of the currently ongoing and future galaxy surveys aim to map significant fractions of the sky and therefore observe a much greater number of high redshift targets.

Additionally, one of the primary goals of future surveys is to determine the cosmic expansion history of the Universe and determine whether the dark energy equation of state parameter deviates from $w_{\rm DE} = -1$, the value assumed in the ΛCDM cosmology. Achieving this goal will require robust distance-redshift measurements to be made over a large baseline in redshift. One possible way of doing this is to use the *Baryon Acoustic Oscillation* (BAO) peak in the clustering signal of galaxies as a standard ruler, with which estimates of the angular diameter distance, as a function of redshift, can be made (e.g. Blake & Glazebrook, 2003). Since, however, the BAO peak appears in the galaxy clustering signal at a scale of ~ 150Mpc, it is necessary for future galaxy surveys to have solid angles that are much larger than the current high redshift pencil beam surveys. For instance, despite Eisenstein et al. (2005) making a detection of the BAO peak in the correlation function of luminous red galaxies (LRGs) in the SDSS, their detection was

rather noisy due to the size of the SDSS sample used and the large scales that need to be probed. (Note that in this case, the problem was the size of the redshift interval covered, rather than the solid angle). The demand, therefore, is for large galaxy surveys extending over a wide redshift interval and covering a significant fraction of the sky.

Given these requirements the main problem now facing galaxy surveys is the accurate measurement of the position of hundreds of thousands, towards millions, of galaxies. The difficulty of obtaining spectroscopic measurements for faint galaxies at high redshift means that many future galaxy surveys are being forced to use photometric techniques to estimate galaxy redshifts. A photometric estimate of a galaxy redshift is made by measuring the flux of the galaxy in multiple wavebands to provide a crude sampling of the spectral energy distribution (SED) of the galaxy. Broad features in galaxy SEDs, such as the 4000Å break, lead to populations occupying redshift-dependent regions in a multi-dimensional colour-space. Measuring the photometry of a galaxy and locating it in this colour-space can therefore provide an estimate of its redshift (e.g. Loh & Spillar, 1986; Connolly et al., 1995). Although many techniques exist to automate this process (e.g. Bolzonella et al., 2000; Benítez, 2000; Arnouts et al., 2002; Collister & Lahav, 2004; Feldmann et al., 2006; Brammer et al., 2008), they are not perfect and still incur large uncertainties on the redshift estimates (typically $3 - 10$ per cent, compared to typical spectroscopic uncertainties of ~ 0.1 per cent). This error is set by the use of broad band filters which smooth out spectral features. Furthermore, in photometric redshift catalogues a fraction of galaxies will be assigned completely the wrong redshift. These *catastrophic failures* occur when a filter is assumed to sample an incorrect feature in the spectra of galaxies at vastly different redshifts (e.g. confusing the Lyman break and the 4000Å break), leading to these galaxies having an incorrect colour and, therefore, an incorrect photometric estimate of their redshift.

The large uncertainties on the photometrically estimated redshifts of galaxies complicates our measurement of galaxy clustering statistics. Calculating the galaxy correlation function at a particular epoch requires selection of galaxies in a narrow redshift window around this epoch. Large uncertainties in galaxy redshift measurements mean that such a sample of galaxies is both incomplete, due to galaxies being shifted out of the region, and, equally, contaminated by galaxies at higher or lower redshift being shifted into this region. The result is that the large-scale structure of galaxies within this region is modified and our estimates of the clustering of the galaxies become distorted. This has to be taken into account when comparing observational estimates of clustering from photometric surveys

to theoretical models.

In this chapter we consider the clustering of samples of galaxies identified in a lightcone mock catalogue using observational selections typical of the next generation of galaxy surveys.

The ultimate goal of this work is to, firstly, present for the first time the GALFORM predictions of luminosity dependent clustering in real and redshift space at high redshift and, secondly, to examine the impact on these predictions of measurement in finite redshift slices (to mimic the measurements that could be made from future galaxy surveys). We measure the angular correlation function of galaxies in a lightcone mock catalogue and examine how this procedure affects the recovery of clustering trends with galaxy luminosity.

We first, in Section 6.1, introduce the definition of the correlation function before discussing sample selection in Section 6.2. In Section 6.3 we present the model predictions at a fixed snapshot. These predictions are compared with the angular clustering in the lightcone in Section 6.4. In Section 6.5 we examine how the peculiar velocities of galaxies affect the estimates of angular clustering. Finally, we summarise and discuss future work in Section 6.6.

6.1 The galaxy angular correlation function

A statistic commonly used to quantify the clustering of galaxies is the two-point galaxy correlation function, $\xi(r)$, which describes the excess probability, $\mathrm{d}P$, of finding a pair of galaxies at a given separation, r, relative to that expected for a random distribution,

$$\mathrm{d}P = n_0^2 \left[1 + \xi(r)\right] \mathrm{d}V_1 \mathrm{d}V_2, \tag{6.1}$$

where n_0 is the number density of the population of galaxies and $\mathrm{d}V_1$ and $\mathrm{d}V_2$ are the volume elements inside which each of the pair of galaxies is found (Peebles, 1980).

Unfortunately, datasets with large errors in the radial distance measurements will lead to noisy and biased estimates of galaxy clustering. Galaxy surveys for which galaxy redshifts have been measured photometrically are a typical example of such a dataset. However, analogous to the 3-dimensional two-point correlation function, is the angular correlation function, $\omega(\theta)$, which simply describes the excess probability, $\mathrm{d}P$, of finding two objects separated by an angle, θ, on the sky, i.e.

$$\mathrm{d}P = \varsigma^2 \left[1 + \omega(\theta)\right] \mathrm{d}\Omega_1 \mathrm{d}\Omega_2, \tag{6.2}$$

where ς is the surface density of objects and $\mathrm{d}\Omega_1$ and $\mathrm{d}\Omega_2$ are the solid angle elements inside which each of the pair of objects resides. By measuring the clustering of galaxies using only their angular separation, the angular correlation function has the advantage that it requires only the 2-dimensional information for the galaxy positions and can be applied to catalogues of galaxies with large redshift uncertainties. Additionally, by using direct observables and hence not requiring co-moving distances to the galaxies, the angular correlation function is a cosmology-independent measurement that can be calculated rapidly, even for galaxy surveys with a complex footprint on the sky. For this reason, it provides a convenient measure of clustering allowing a comparison between both theoretical predictions and observational datasets.

The angular correlation function is calculated by simply generating a set of randomly distributed points with the same footprint on the sky as the real dataset and then counting the number of data-data, $\mathrm{DD}(\theta)$, random-random, $\mathrm{RR}(\theta)$, and data-random, $\mathrm{DR}(\theta)$ pairs in bins of angular separation, $\theta \rightarrow \theta + \mathrm{d}\theta$. Counting the number of random-random pairs provides an estimate of the mean density of the dataset so that data-data pair counts will reveal correctly calibrated density fluctuations. Several clustering estimators exist to then translate these pair counts into a value for $\omega(\theta)$ (e.g. Davis & Peebles, 1983; Hamilton, 1993; Landy & Szalay, 1993). A simple first-order estimate for $\omega(\theta)$ can be made by calculating the fractional enhancement in the number of data-data pairs above random, i.e.

$$\omega(\theta) = \frac{DD(\theta)}{RR(\theta)} - 1. \tag{6.3}$$

This (and other linear estimators) can often provide biased estimates due to errors in the mean density resulting from small sample sizes or interaction with the edges of the solid angle footprint. A widely used estimator that is second-order in errors in the mean density, the error distribution for which is Poissonian, is the Landy-Szalay estimator,

$$\omega(\theta) = 1 + \left(\frac{N_\mathrm{r}}{N_\mathrm{d}}\right)^2 \frac{\mathrm{DD}(\theta)}{\mathrm{RR}(\theta)} - 2\left(\frac{N_\mathrm{r}}{N_\mathrm{d}}\right) \frac{\mathrm{DR}(\theta)}{\mathrm{RR}(\theta)}, \tag{6.4}$$

where N_d and N_r are the number of data and random points respectively (Landy & Szalay, 1993). A determination of $\omega(\theta)$ with an arbitrary accuracy and reduced statistical noise, especially at small separations, can in principle be achieved by either using a small number of realisations of high density random points (e.g. 1-2 realisations with $N_\mathrm{r} \sim 10N_\mathrm{d}$) or by using a large number of realisations of low density random points (e.g. 10 realisations with $N_\mathrm{r} \sim N_\mathrm{d}$). Since the computation time required to calculate the separation between N objects scales naively as N^2, then the latter approach is preferable.

The angular correlation function has been shown to be fitted well by a power law and is typically expressed as,

$$\omega(\theta) = A_\omega \theta^{-\gamma}, \tag{6.5}$$

with normalisation, A_ω, and power-law slope $\gamma \approx 0.8$ (e.g. Groth & Peebles, 1977; Roche & Eales, 1999). However, since the total pair counts summed over all bins is fixed (equal to $\frac{1}{2}n(n-1)$), the positive clustering signal observed at small angular separations demands that $\omega(\theta)$ becomes negative at large scales. When fitting $\omega(\theta)$ with a power-law, this is obviously a problem. Worse still is that for surveys with sufficiently small sky coverage, the value of $\omega(\theta)$ remains positive, even on large scales. This can be the case when the footprint of such galaxy surveys hits a galaxy cluster, leading to the mean galaxy density in the field being biased higher than the true mean density. Since the pair count cannot increase in all angular bins, whilst maintaining a constant number of pairs, the normalisation must be corrected at large scales. This correction, known colloquially as the *integral constraint*, W, relates the observed estimate of $\omega_{\rm est}(\theta)$ to the true value, $\omega(\theta)$, according to,

$$\omega(\theta) \approx \omega_{\rm est}(\theta) + W. \tag{6.6}$$

Assuming the power-law relation from Eq. (6.5), we then have,

$$\omega(\theta) = A_\omega \left(\theta^{-\gamma} - C\right), \tag{6.7}$$

where C is calculated by summing up the random pair counts over the N bins of angular separation,

$$C = \frac{\sum_{i=1}^{N} RR(\theta)\theta^{-\gamma}}{\sum_{i=1}^{N} RR(\theta)}. \tag{6.8}$$

The angular correlation function can be related to the real-space two-point correlation function through the relation (e.g. Peebles, 1980),

$$\omega(\theta) \approx \int_0^\infty {\rm d}r_1 \int_0^\infty {\rm d}r_2 p_1\left(r_1\right) p_2\left(r_2\right) \xi\left(R, \frac{r_1 + r_2}{2}\right), \tag{6.9}$$

where

$$R \equiv \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos\theta}, \tag{6.10}$$

r_1 and r_2 are the co-moving distances to each of a pair of galaxies, and p_1 and p_2 are selection functions projecting the 3-dimensional density contrast, $\delta(r\mathbf{\Theta}, r)$, to the density contrast on the sky, $\hat{\delta}(\mathbf{\Theta})$:

$$\hat{\delta}_{1,2}(\mathbf{\Theta}) = \int_0^\infty {\rm d}r\, p_{1,2}(r)\delta_{1,2}(r\mathbf{\Theta}, r), \tag{6.11}$$

where r is the co-moving radial distance along the line-of-sight, $\mathbf{\Theta}$. The relation in Eq. (6.9), which is valid out to large angular separations, can be approximated by,

$$\omega(\theta) = \int_0^{\infty} \mathrm{d}\bar{r} p_1(\bar{r}) p_2(\bar{r}) \int_{-\infty}^{+\infty} \mathrm{d}\Delta r \xi\left(\hat{R}, \bar{r}\right), \tag{6.12}$$

where $\hat{R} \equiv \sqrt{\bar{r}^2\theta^2 + \Delta r^2}$, $\bar{r} \equiv (r_1 + r_2)/2$ and $\Delta r \equiv r_2 - r_1$. This approximation is typically referred to as the Limber equation (Limber, 1953), which uses the small angle approximation and the approximation that the correlation length is much smaller than the depth of the galaxy survey to separate the integrals. Recently, Simon (2007) reviewed the accuracy of the Limber equation and concluded that an error of approximately 10 per cent is reached for angular separations larger than a few degrees.

6.2 Sample selection

A crucial step in the calculation of galaxy clustering is the selection of a sample of objects from which to estimate the clustering.

In this work we use a lightcone galaxy mock catalogue, covering ~ 100 square degrees, with galaxies selected according to the selection limits $(\mathrm{g, r, i, z, y}) =$(26.22, 25.86, 25.82, 25.19, 23.75), which have been chosen to match the flux selection criteria of the Pan-STARRS Medium Deep Survey (Kaiser et al., 2002). Galaxies had to be selected in at least one of these five bands to be included in the lightcone. We centre our redshift slice on $z \simeq 0.59$, which lies directly between two of the *Millennium Simulation* snapshots, positioned at redshifts of $z \simeq 0.56$ and $z \simeq 0.62$. This redshift corresponds approximately to the peak in the redshift distribution of our lightcone catalogue (e.g. Cai et al., 2009).

We choose to select galaxies according to their apparent magnitude. Our aim is to attempt to reproduce the steps that would be taken if the lightcone was an observed dataset from a photometric redshift survey. Figure 6.1 shows the apparent magnitude bins adopted and the number density of the galaxies selected in each bin as a function of rest-frame absolute magnitude. To convert between the observer-frame apparent magnitude, m_{AB}, of the galaxy and its rest-frame absolute magnitude, M_{AB}, we have used,

$$m_{\mathrm{AB}} = M_{\mathrm{AB}} + 5\log_{10}\left(\frac{d_L(z)}{10\mathrm{pc}}\right) - 2.5\log_{10}(1+z) + k. \tag{6.13}$$

where z is the redshift of the galaxy, $d_L(z)$ is the luminosity distance to the galaxy and k is a k-correction, which is estimated by `GALFORM`, that is dependent upon the galaxy type. For comparison, we also show in Figure 6.1 the r-band galaxy luminosity function

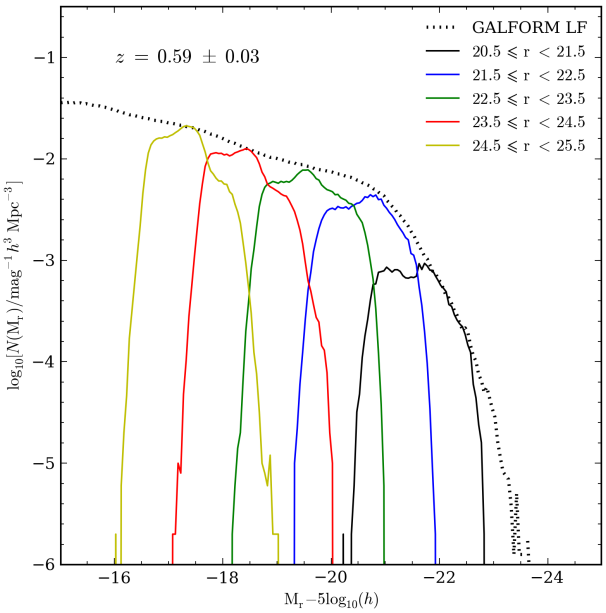

Figure 6.1: The absolute magnitude distribution for galaxy samples selected in apparent magnitude bins, as stated in the legend. Galaxy samples were extracted from lightcone mock catalogue in the redshift range $z = 0.59 \pm 0.03$. The dotted line shows the galaxy luminosity function for galaxies in the *Millennium Simulation* snapshot corresponding to $z \simeq 0.62$.

for the *Millennium Simulation* snapshot corresponding to a redshift $z_{\rm snap} \simeq 0.62$. Due to the way that the lightcone is constructed, this is the snapshot that will contribute the galaxies within the range $0.56 \lesssim z \lesssim 0.62$ (see Chapter 3). The width of this redshift slice is comparable to typical uncertainties on photometric redshift measurements.

6.3 Snapshot two-point correlation functions

In this section we will look at the clustering predictions in an intermediate redshift snapshot. First, because we are selecting galaxies in apparent magnitude bins from the lightcone, we need to see the implications this has for the selection of galaxies in absolute magnitude in the snapshot.

We calculate the angular correlation function for the various apparent magnitude bins using a redshift window $0.56 \lesssim z \lesssim 0.62$, i.e. constrained by the redshifts of the two adjacent snapshots, and compare the trend in the angular clustering signal with the trend in the 3-dimensional real-space and redshift space clustering signals for similarly selected galaxy samples in the $z \simeq 0.62$ snapshot. Note that the width of this redshift slice is comparable to the typical photometric redshift uncertainty with broad band photometry.

We compare this to the $z \simeq 0.62$ snapshot since, due to the construction method for the lightcone, this is the snapshot that will provide the galaxies within the redshift range $0.56 \lesssim z \lesssim 0.62$. Additionally, we do this to check whether any clustering trend seen in the snapshot is preserved in the construction of the lightcone and prior to our introduction of any redshift error (and, therefore, contamination from galaxies at higher and lower redshifts, which will have come originally from alternative snapshots in the simulation and which may have slightly different dependencies of clustering strength on luminosity).

The first step in the calculation of the real-space and redshift-space 3-dimensional correlation functions is to ensure that we are selecting a comparable sample of galaxies in the snapshot as we are in the lightcone. Due to different galaxy types displaying different k-corrections, and the finite width of the redshift slice, a sample of galaxies selected using a top-hat function in apparent magnitude will not display a top-hat distribution of absolute magnitudes. Instead, the galaxy dependent k-correction and the finite width of the redshift slice will increase the width of the absolute magnitude bin, which will also be sampled non-homogeneously.

When comparing to the luminosity dependent clustering in the simulation snapshot, we need to ensure that we are sampling the same distribution of absolute magnitudes as seen in the lightcone redshift slice. One can use the distribution of the absolute magnitudes of the galaxies selected in the lightcone to construct a probability distribution that is a function of absolute magnitude. We do this by comparing the number density of objects in the lightcone, within a given absolute magnitude, to the number density of objects in the snapshot within the same absolute magnitude range. That is, we divide the absolute magnitude distribution of the galaxies in the lightcone by the luminosity function of the galaxies in the snapshot of interest. This provides us with the necessary probability distribution as a function of absolute magnitude in order to select galaxies with the same range of magnitudes that are sampled in the lightcone. Note that we have a separate probability distribution for each of the apparent magnitude samples shown in Figure 6.1. For each galaxy in the snapshot, we can interpolate over the probability distribution for each apparent magnitude sample to obtain the probability as to whether that galaxy would be selected in that particular sample. By comparing this result to a randomly generated number we either select or reject the galaxy.

The real-space and redshift-space 3-dimensional correlation functions computed in this way for the $z \simeq 0.62$ snapshot of the *Millennium Simulation* are shown in Figure 6.2. In these figures, the galaxy samples are labelled according to the apparent magnitude

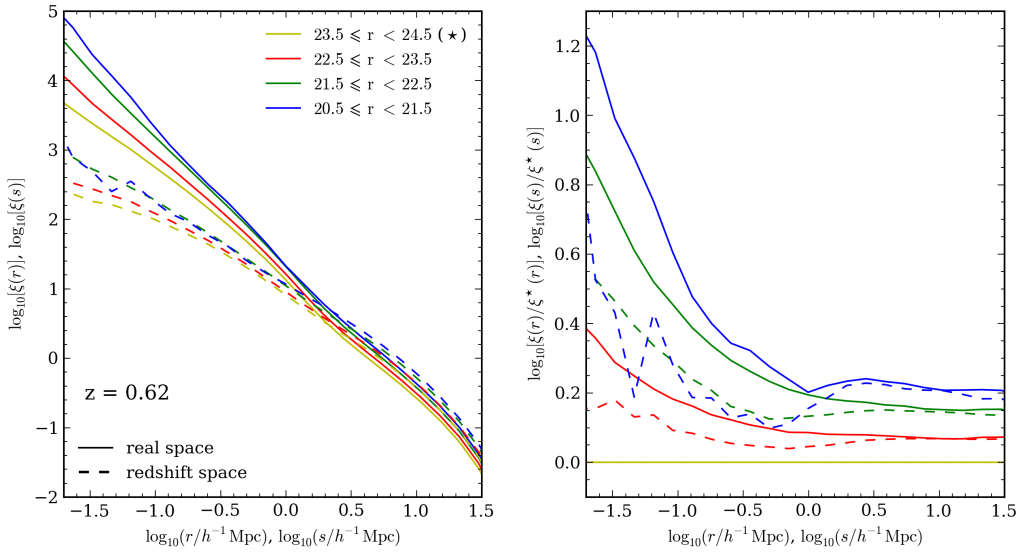

Figure 6.2: Left: Real-space and redshift-space correlation functions for galaxies in the $z \simeq 0.62$ snapshot of the *Millennium Simulation* selected according to the apparent magnitude cuts applied to the lightcone catalogue. Right: Correlations, as plotted in the left panel, now plotted relative to the clustering amplitude of the faintest sample (indicated by the star) in either real-space or redshift-space, depending on the ratio being plotted. In both panels, the real-space correlation function is plotted using solid lines, whilst the redshift-space correlation function is plotted using dashed lines.

samples to which they correspond.

Examining first the 3-dimensional real-space correlation function, we see that there exists a clear trend that more luminous galaxies are more strongly clustered, i.e. the clustering amplitude is larger for the samples typically containing more luminous galaxies in agreement with results in the literature at $z = 0$ (e.g. Norberg et al., 2002; Zehavi et al., 2005, 2011). This trend is visible at all scales considered.

To calculate the redshift space correlation function we use the distant observer approximation to transform the real-space Cartesian X-coordinate of the selected galaxies according to

$$X' = X + \frac{v_X \, (1 + z_{\rm snap})}{H \, (z_{\rm snap})}, \tag{6.14}$$

where X' is the redshift space X-coordinate of the galaxy, v_X is the peculiar velocity of the galaxy, in $\mathrm{km\,s^{-1}}$, along the X-axis and $H\,(z_{\rm snap})$ is the Hubble parameter at the redshift of the snapshot, $z_{\rm snap}$. We can see that the dominant effect of redshift-space distortions is to suppress the clustering strength at small scales. The suppression of the clustering strength is greater for more luminous galaxies. There is a smaller boost in amplitude on large scales on moving to redshift-space. The trend of increasing clustering

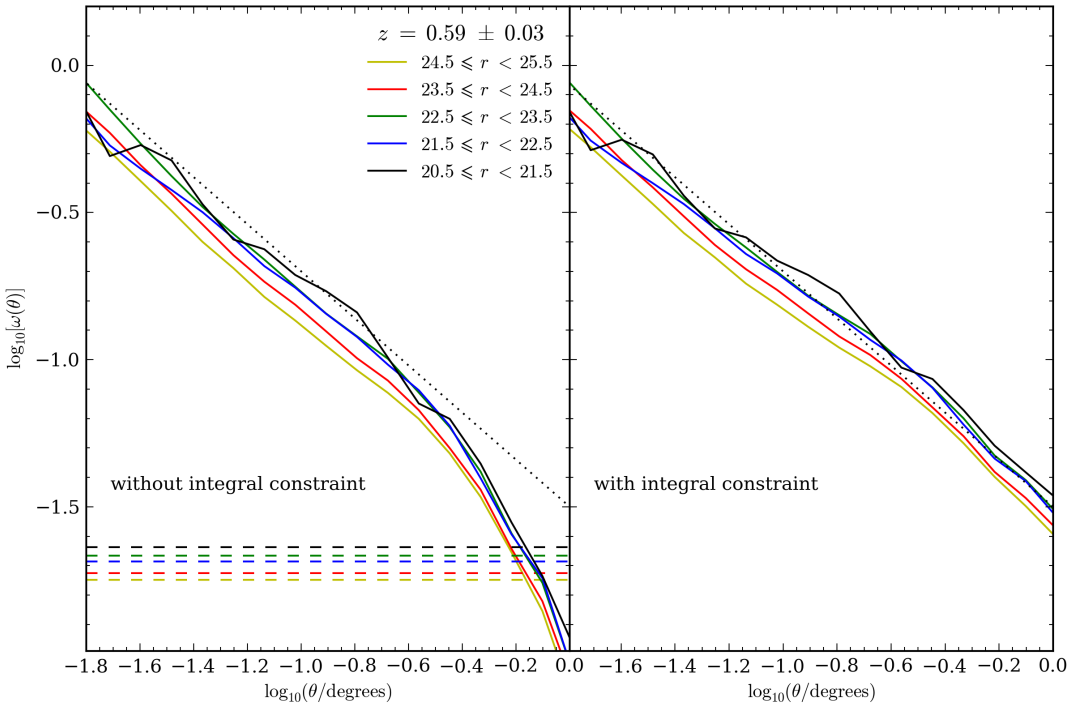

Figure 6.3: Angular correlation function for apparent magnitude selected galaxy samples, extracted from a redshift slice $z = 0.59 \pm 0.03$ in the mock galaxy lightcone, plotted both without the integral constraint (left) and with the integral constraint (right). In both panels, the dotted line corresponds to a power-law with slope -0.8. In the left-hand panel, the horizontal dashed lines indicate the value of the integral constraint for each apparent magnitude sample.

strength with increasing galaxy luminosity is preserved at large scales (above $1\,h^{-1}$Mpc).

6.4 Angular clustering in galaxy lightcone

We now consider whether the same trend with luminosity that was seen in the snapshot in Figure 6.2 is present in the angular clustering in the galaxy lightcone mock catalogue. In Figure 6.3 we show the angular correlation function for apparent magnitude selected galaxy samples, extracted from within the redshift range, $0.56 \leqslant z \leqslant 0.62$. We have neglected peculiar velocities and have assumed a zero redshift measurement error. This redshift window has been chosen since it spans the redshift interval between two of the snapshots in the *Millennium Simulation*. In Figure 6.3 we show the angular correlation function both before and after being corrected by the integral constraint. Here we estimate the integral constraint using the power-law relation, shown in Eq. (6.5), normalised at an angle of 0.5°. We can clearly see the effect of introducing the integral constraint at large

Table 6.1: The number of galaxies in each of the apparent magnitude bins in the redshift range $z = 0.59 \pm 0.03$ of the lightcone catalogue.

Apparent magnitude limits	Number of galaxies
$24.5 \leqslant \mathrm{r} < 25.5$	23,264
$23.5 \leqslant \mathrm{r} < 24.5$	15,782
$22.5 \leqslant \mathrm{r} < 23.5$	10,611
$21.5 \leqslant \mathrm{r} < 22.5$	5,683
$20.5 \leqslant \mathrm{r} < 21.5$	1,231

angular scales.

The trend in the clustering strength with luminosity is much less clear in the lightcone than in the simulation snapshot. Although the two faintest samples considered appear to show a positive correlation between the luminosity and the clustering strength, the brightest three samples are much noisier and display less of a trend. The brightest sample considered does show the strongest clustering amplitude, but it is very noisy due to the limited size of the sample, which is indicated in Table 6.1. This is shown more clearly in Figure 6.4 where we plot the clustering strength of each sample, relative to the strength of the faintest sample. Error bars indicate the size of the error on the ratio of $\omega(\theta)$ given the Poisson errors on the pair counts for the two apparent magnitude samples being compared. Any trend in clustering strength with luminosity is difficult to see due to the next two brightest samples ($21.5 \leqslant \mathrm{r} < 22.5$ and $22.5 \leqslant \mathrm{r} < 23.5$), which, for all angular separations above $\sim 0.1^\circ$, have similar clustering strengths but which are also noisy. The clustering of the $21.5 \leqslant \mathrm{r} < 22.5$ sample decreases at the smallest angular scales shown. The brightest bin appears to show a similar decrease in clustering amplitude for small angles, though this is not certain given the noise on this estimate. Therefore, the two faintest apparent magnitude bins considered suggest that the clustering strength increases with luminosity, though the levels of noise in the three brightest bins make it unclear whether the trend extends to brighter luminosities.

To further examine the trend in the angular clustering from the lightcone, we plot in Figure 6.5 the clustering amplitude for galaxies in the four brightest apparent magnitude bins according to the real-space, redshift-space and angular correlation functions. To convert the angular separations into co-moving transverse separations we have, for the *Millennium Simulation*, determined the angular scale at the redshift at the midpoint of

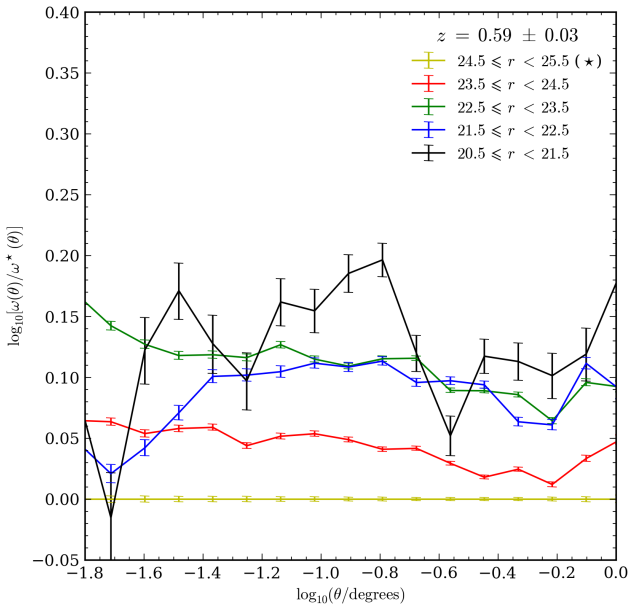

Figure 6.4: Comparison of the amplitude of the angular correlation function for galaxy samples selected in increasingly bright apparent magnitude bins. All samples are plotted relative to the clustering amplitude of the faintest sample. Error bars indicate the size of the error on the ratio of $\omega(\theta)$ given the Poisson errors on the pair counts for the two apparent magnitude samples being compared.

the redshift bin. As we have commented above, there is a clear sequence of clustering amplitudes varying with luminosity in real-space and redshift-space. This trend disappears for the brightest two samples when considering the angular clustering. In future work we will investigate whether the angular clustering is better described as an integral over $\xi(r)$, $\xi(s)$ or some combination in a thin redshift slice such as the one considered here. One possibility is that a large galaxy cluster is present in the lightcone within the redshift range of interest. We have calculated the angular correlation function for two separate lightcones where the observer was placed in the same location, but with a different orientation. The discrepancy in the trend between the angular and spatial clustering is visible in the results for both lightcones. As a further check, we have calculated the angular correlation function using both the Hamilton (1993) and Landy & Szalay (1993) estimators, both of which lead to clustering signals that display the same discrepancy.

6.5 Introducing peculiar velocities

So far, we have estimated the angular correlation function for galaxies based upon the cosmological redshift of the galaxies in the lightcone. We now briefly consider the redshift

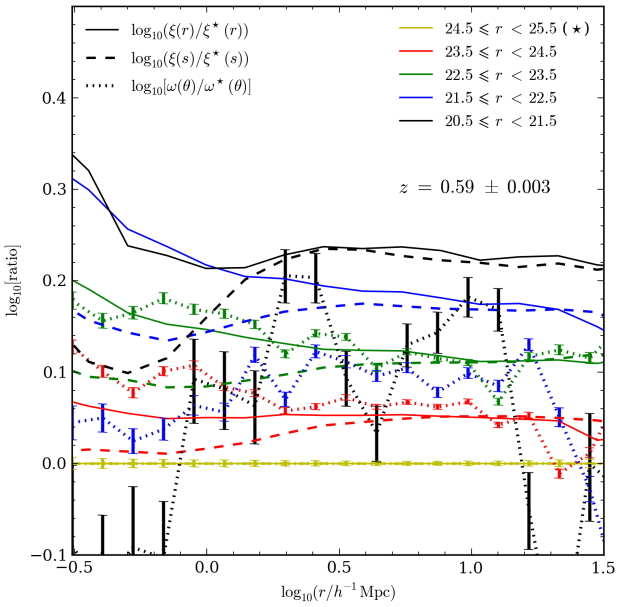

Figure 6.5: Comparison of the clustering amplitude for the four brightest apparent magnitude galaxy samples in the redshift range $z = 0.59 \pm 0.03$. The ratio of the real-space clustering is shown by the solid lines, the ratio of the redshift-space clustering is shown by the dashed lines and the ratio of the angular clustering is shown by the dotted lines. The clustering strengths of each correlation function are plotted relative to the clustering strength in the faintest bin (indicated with a star) for that particular correlation function. Error bars indicate the size of the error on the ratio of $\omega(\theta)$ given the Poisson errors on the pair counts for the two apparent magnitude samples being compared.

each galaxy has in the lightcone when the peculiar velocities are included and the effect that the peculiar velocities have on the clustering signal. The redshift slice we use is kept fixed at $z = 0.59 \pm 0.03$. We note that by including the peculiar velocities of galaxies in their redshifts, the change in the redshift of a galaxy has an rms value of $\delta_z \simeq 0.001$, which is much smaller than the redshift interval we consider.

In Figure 6.6, we show the angular correlation function for the same apparent magnitude samples, as adopted in Figure 6.3. Although the trend in the clustering amplitude with luminosity is the same as before, the overall amplitude of the clustering signals are higher by 0.1 dex, especially towards large angular separations. The change in the amplitude of $\omega(\theta)$ when the peculiar velocities of galaxies are included in their redshifts, shows that for thin redshift slices the angular clustering is not necessarily proportional to the integral of the real-space spatial correlation function, but might in fact be proportional to the redshift-space spatial correlation function, or a combination of the two.

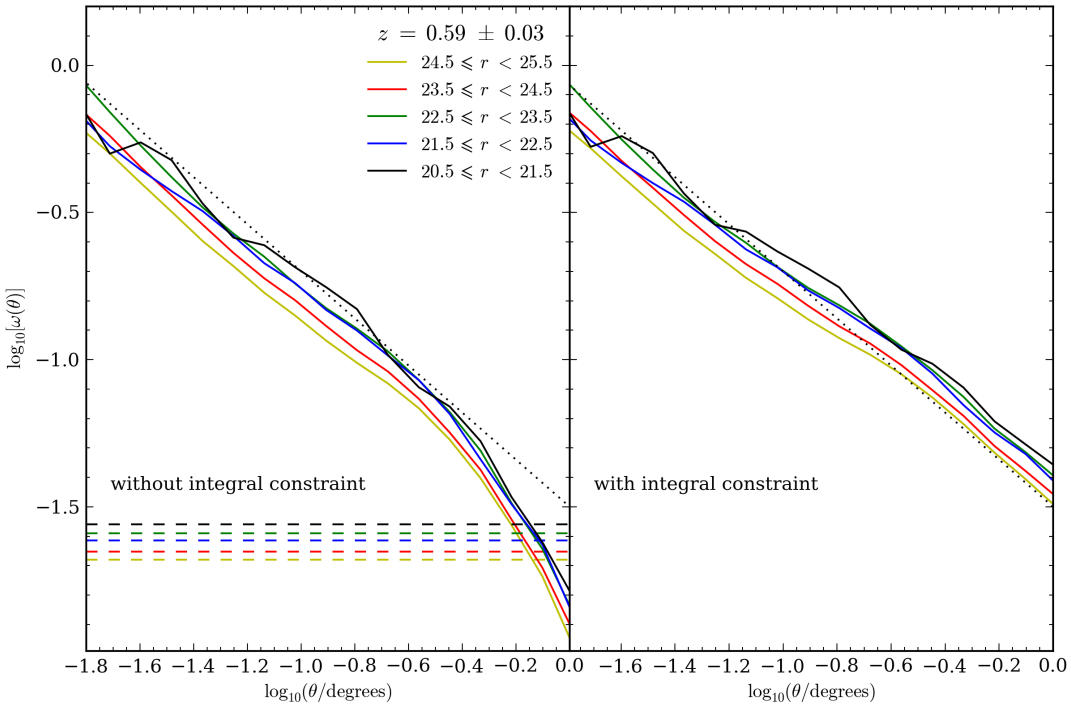

Figure 6.6: The angular correlation function of galaxies in apparent magnitude bins obtained when the peculiar velocities of galaxies are included in their redshifts. The various lines have the same meaning as in Figure 6.3.

6.6 Conclusions & future work

In this preliminary analysis, we have used the lightcone mock catalogues developed in this thesis to examine the variation in the angular clustering strength of galaxy samples selected according to different apparent magnitudes. A clear trend in the clustering strength with luminosity is observed in the real-space and redshift-space correlation functions for galaxies extracted from a simulation snapshot ($z \simeq 0.62$) using bins of apparent magnitude. For galaxies selected from a lightcone mock catalogue using the same apparent magnitude bins, the amplitude of the angular correlation function shows a less clear trend with luminosity. The two faintest apparent magnitude bins suggest the presence of a trend, but the three brightest bins have angular clustering amplitudes that are very similar. It is not clear, therefore, whether the angular clustering amplitude shows any trend with luminosity. If the trend is present then our results suggest that the trend is much weaker for brighter galaxies. The trend could perhaps be made clearer by increasing the size of the galaxy samples, listed in Table 6.1. This could be done by either increasing the width of the redshift shell or increasing the solid angle of the lightcone.

The further work of this project is:

- Examine the angular correlation function for many more lightcone catalogues, constructed using alternative galaxy formation models to see whether the discrepancy in the angular clustering amplitude is model-dependent or being affected by the presence of large-scale structure. The effects of cosmic variance could also be assessed by comparing estimates of $\omega(\theta)$ for lightcones with larger solid angles.

- Assess whether the angular clustering of galaxies in redshift slices is probing the real-space or redshift-space correlation functions, or a combination of the two, and, therefore, what is the appropriate correlation function to use in the Limber equation. Furthermore, we must determine whether the Limber equation holds for such thin redshift slices.

- Introduce a simple Gaussian redshift measurement error to examine how the ordering of the amplitude of $\omega(\theta)$ with luminosity is affected by uncertainties in the redshift measurement, leading to interlopers in the redshift range of interest. If the trend is found to be present, it is therefore interesting to determine whether the trend is preserved with increasing redshift measurement error and what error can be tolerated before the trend is washed out. Furthermore, we would need to introduce redshift errors that are dependent upon galaxy type and, therefore, a more realistic representation of photometric redshift errors. This could be done by using an off-the-shelf code to determine a photometric redshift, and associated measurement error, for each galaxy in the lightcone. Using a photometric redshift estimator would provide an uncertainty that is dependent upon galaxy type as well as realistic effects such as catastrophic errors, which would dilute the clustering signal.

- Consider whether any trend in the angular clustering with luminosity is present when selecting galaxies using narrow band filters that have been used by the ALHAMBRA survey (Moles et al., 2008) and will be used by the Physics of the Accelerating Universe (PAU) survey (Benítez et al., 2009).

- Investigate the effect of changing the size and number of apparent magnitude bins used to select the galaxies and whether the discrepancy is dependent upon either of these properties.

- Further select the galaxies by colour and examine whether (i) the luminosity-dependent angular clustering displays a clear trend for red or blue galaxies and

(ii) whether the angular clustering amplitude displays any dependence on colour alone (i.e. irrespective of the brightness of the galaxy).

- Estimate $\omega(\theta)$ for thin radial slices centred on different redshifts, to examine whether any trend in the clustering amplitude with luminosity or colour evolves with redshift.

- The angular correlation function for flux limited samples is often scaled to a common depth. Doing this allows one to test for systematics (e.g. Maddox et al., 1990). We plan to consider such a rescaling to investigate the extent to which systematics are affecting our estimates of the angular clustering for flux limited samples in thin redshift slices.

Chapter 7

Conclusions & future work

In this thesis we have focused on the construction of mock galaxy catalogues from the GALFORM semi-analytical model of galaxy formation and considered several applications of these mock catalogues for current and future galaxy surveys.

7.1 Construction of galaxy lightcone mock catalogues

We have presented, for the first time, a method for constructing end-to-end mock galaxy lightcone catalogues from the GALFORM semi-analytical model of galaxy formation, applied to the halo merger trees extracted from a cosmological N-body simulation.

The mock catalogues that we construct are *lightcone* catalogues, in which a galaxy is placed according to the epoch at which it first enters the past lightcone of the observer. We use interpolation to determine the positions of galaxies at epochs intermediate to the simulation snapshots, which represents an improvement over previous work. We have shown that our adopted interpolation scheme leads to accurate predictions for real space galaxy clustering down to scales well within the one-halo regime. This is the first lightcone construction method that uses interpolation in this way.

We do not attempt, at this time, to interpolate the other properties of galaxies between the snapshot outputs due to the complex time evolution of many of the galaxy properties. We note, however, that by simply adopting the properties a galaxy had immediately before it entered the lightcone, we still incorporate the evolution of galaxy properties with cosmic time. We do, however, use interpolation to smooth out the k-correction in the galaxy magnitudes. This means that we are using a consistent definition of the observer-frame magnitude for the redshift at which the galaxy appears in the lightcone. Note, however, that this interpolation simply accounts for this redshift dependent effect and no attempt is made to evolutionary correct the magnitudes between the snapshot outputs.

Our approach has a number of attractive features. First, we use a physical model of galaxy formation which makes *ab initio* predictions. This means that we can build mocks for epochs or selections which are currently unprobed. Empirical approaches, such as HOD modelling, are not able to do this, as they depend on the existence of observations for their calibration. Second, our construction method is generic and is not tied to a particular choice of N-body simulation or semi-analytic model. As better N-body simulations or more accurate galaxy formation models become available our method can still be used. Our scheme could be applied to the output of other semi-analytical models that have been used to populate the output of an N-body simulation. Third, the semi-analytic model that we have used has a unique multi-wavelength capability, which means that we can mimic a wide variety of galaxy surveys spanning a large range of wavelengths.

Future developments to the method beyond what we have presented here include:

- Implementation of a method to interpolate the evolution of other galaxy properties, such as stellar mass, between simulation snapshots. One of the main difficulties is the complex time evolution of these properties and how to interpolate these properties in the event of a galaxy merger.

- Extend the `GALFORM` model to output spectra for all or selected samples of galaxies according to their location in a particular galaxy lightcone.

- Include a method for applying selection masks to the lightcones to enable them to provide even more realistic predictions for galaxy surveys.

- Incorporate a post-processing pipeline for constructing mock galaxy images.

- Introduce lensing effects into the lightcones. Doing this will require ray tracing of the light and so would more than likely need to be carried out in post-processing. Additional to this is the inclusion of phenomena such as intrinsic alignments, which will require an improvement in how the `GALFORM` model determines the angular momentum of the dark matter halos and how the angular momentum of the hosted galaxies couples to the angular momentum of the halo.

- Developing an efficient method of distributing lightcone mock catalogues to the wider community. The huge number of galaxies that will be observed in future galaxy surveys and the extensive list of galaxy properties that can be predicted for each galaxy will mean that mock catalogues emulating such galaxy surveys will be

incredibly large, of the order several hundred gigabytes, even terabytes, of data. One possible approach would be to construct a SQL database to serve catalogues for many different galaxy surveys and allow users to further select galaxies using SQL queries.

- Construct lightcones by using the `GALFORM` model in conjunction with other cosmological N-body simulations of much larger volume or that use alternative cosmological models. Lightcones constructed from such simulations will be vital for making clustering predictions for future galaxy surveys that aim to test the ΛCDM paradigm.

7.2 Applications of mock catalogues

Following on from this, we have focused on three example applications of these mock galaxy catalogues, emphasizing their use as essential tools for calibration of techniques/estimators and as a predictive tool for ΛCDM. These applications all clearly demonstrate the main advantage of mock catalogues; that in any circumstance we know the "true" value of a statistic that we are be aiming to recover using a statistical estimator for a galaxy survey.

7.2.1 Effectiveness of the BzK colour selection technique

In our first application we consider the effectiveness of the BzK colour selection technique, which was designed to isolate galaxies in the redshift interval $1.4 < z < 2.5$. The predictions that we present here are the first investigation of the BzK selection technique using the `GALFORM` model.

The lightcone mock catalogue that we adopt is in reasonable agreement with the observed number counts of all BzK galaxies, as well as with the observed counts of the subsample of BzKs that are star-forming galaxies. We predict that over 75 per cent of the model galaxies with $K_{AB} \leqslant 23$, and $1.4 < z < 2.5$, are selected by the BzK technique. Interloper galaxies, outside the intended redshift range, are predicted to dominate bright samples of BzK galaxies (i.e. those with $K_{AB} \leqslant 21$). Fainter K-band cuts are necessary to reduce the predicted interloper fraction. We also show that shallow B-band photometry can lead to confusion in classifying BzK galaxies as being star-forming or passively evolving.

Overall, we conclude that the BzK colour selection technique is capable of providing a sample of galaxies that is representative of the $1.4 < z < 2.5$ galaxy population.

Proposed future work to extend this project includes:

- Considering how the effectiveness of the BzK selection affects the measurement of the angular clustering of galaxies at $z \sim 2$. Using a lightcone mock catalogue, one could calculate the angular correlation function for all BzK selected galaxies and compare this to the angular correlation function of galaxies that are lying in the redshift range $1.5 < z < 2.5$ to see the effect that interloper galaxies have on the clustering signal.

- Use additional mock galaxy catalogues to investigate the effectiveness of colour selection techniques used to identify other populations of galaxies, including luminous red galaxies (LRGs) and extremely red objects (EROs).

- Investigate the BzK number counts in mock catalogues constructed from different variants of the `GALFORM` model.

7.2.2 Calibration of a galaxy group finding algorithm

The second application that we consider is the use of a mock galaxy catalogue to calibrate a galaxy group finding algorithm. The algorithm that we have adopted is the Eke et al. (2004a) friends-of-friends algorithm, which we plan to ultimately apply to the 6-degree Field Galaxy Survey (6dFGS). In this application we have applied a slightly different method for constructing mock catalogues, due in part to the fact that the 6dFGS is a shallow galaxy survey (with median redshift $z_{1/2} \simeq 0.05$) and the characteristics of the group-finding algorithm, which relies upon a percolation technique.

We have vindicated an alternative, completely objective method for calibrating a group-finder based upon the recovery of the distributions of several group properties, including the redshift distribution, the distribution of group luminosities, the distribution of the line-of-sight velocity dispersion of groups and the distribution of the projected radii of groups. These properties were chosen because they are, on the whole, easily measurable and provide constraints on the geometry and membership of groups.

Our calibration has shown that it is extremely difficult to determine the optimal linking lengths by minimising the χ^2 for individual properties. However, we find that by combining the χ^2 statistics for multiple properties, we can identify a set of linking lengths with which we can construct a catalogue of groups that provides a reasonable match to the distribution of halo properties. By combining the χ^2 values for various properties, we find that the linking lengths $(b_\perp, R) = (0.07, 14)$ appear to provide a reasonable

match to the majority of the properties. We find, however, that increasing the minimum richness of the groups one wishes to recover leads to larger values for the optimal linking lengths. Therefore, groups catalogues obtained using different linking lengths are required to faithfully recover the properties of groups with different richness. Similar is true for the groups properties, in that it is not possible to locate a single set of linking lengths that provides optimum recovery of all halo properties, at least for groups with low multiplicity. Which properties one wishes to prioritise, will determine the required values of the linking lengths.

We have discussed several extensions to this work at the end of Chapter 5. We summarise here the immediate extensions.

- Apply the 6dFGS completeness mask to the 6dFGS mock catalogue to ensure that the mock catalogue provides as realistic a resemblance as possible of the real 6dFGS.

- Consider whether any other group properties are necessary to improve the calibration method we have presented here and how to effectively decide upon the optimum linking lengths to adopt, given the dependency on group properties such as richness.

- Perform a one-to-one matching procedure to cross-check that our calibration method is correctly identifying the optimum linking lengths to use (e.g. following Eke et al., 2004a).

- Once calibration is complete, apply the group-finder to the 6dFGS to construct a 6dFGS groups catalogue. Used in combination with the `GALFORM` semi-analytic model and observational datasets at other wavelengths, such a catalogue could be used to examine the content and structure of galaxy groups in the local Universe.

7.2.3 Angular clustering predictions at high redshift

For our third and final application we use a lightcone mock catalogue to make the first predictions from `GALFORM` of the clustering of galaxies at $z \sim 0.6$. This redshift corresponds to the peak in the redshift distribution of galaxies that will be observed in future galaxy surveys, such as the Pan-STARRS survey.

We have measured the clustering of galaxies, selected in bins of apparent magnitude, by estimating the angular correlation function, $\omega(\theta)$, for a thin redshift slice centered on $z \sim 0.6$. We have chosen to estimate $\omega(\theta)$ as it is a cosmology independent quantity that

can be estimated relatively easily for samples of galaxies with large uncertainties on their redshift measurement.

We have compared our estimates of $\omega(\theta)$ with an estimate of the 3-dimensional real-space and redshift-space correlation functions, taken from the $z \simeq 0.62$ snapshot of the *Millennium Simulation*. Before making the comparison, we consider how to properly select galaxies from the simulation snapshot such that they have the same distribution of absolute magnitudes as our apparent magnitude selected galaxies in the lightcone. Although the real-space and redshift-space clustering displays a trend of increasing clustering amplitude with increasing luminosity, this trend is less clear in the $\omega(\theta)$ estimates from the lightcone due to the clustering estimates for the brightest two apparent magnitudes being very noisy. This could point to the need to use a wider redshift bin in our analysis. The original bin width was chosen to be comparable to the typical error on photometric redshifts from broad band photometry.

The work presented in this application, is at a very early stage. We have discussed the necessary extensions to this project at the end of Chapter 6. For brevity, we summarise the immediate extensions to the work below:

- Determine whether the angular correlation function is probing the real-space or redshift-space clustering of galaxies (or a combination of the two) and if the Limber equation is appropriate in the case of narrow redshift slices.

- Examine the trend in clustering amplitude using many different lightcones to check that cosmic variance is not affecting our estimates of the angular correlation function. This could also be done by using lightcones covering larger solid angles. Scaling the estimates of $\omega(\theta)$ to a common depth could help identify other possible systematics.

- Examine the effect on the angular clustering signal of introducing a redshift measurement error. A simple error could be introduced in the first case, before introducing a more realistic error that is, for example, dependent upon galaxy type. The latter could be implemented by carrying out an estimate of the photometric redshifts using an off-the-shelf code.

- Investigate whether there is a trend between the clustering amplitude and galaxy colour and, if so, how redshift measurement error affects this trend.

- Examine how the (possible) trend with luminosity and colour changes if we switch to

using narrow band filters rather than broad band filters, and whether photometric redshift errors from narrow bands preserve the trend down to fainter flux limits.

7.3 Proposals for future work

Extensions of the work presented in this thesis will be geared towards large volume, wide angle surveys, such as Pan-STARRS, the Dark Energy Survey and the EUCLID survey. Some of the issues which need to be addressed in the next generation of mock catalogues includes:

- Current large volume cosmological simulations are limited by resolution. One possible project would be to investigate methods of placing halos from smaller volume simulations, with higher resolution, into the low resolution halos in the larger simulation. For each halo in the larger simulation, one would need to identify a halo in the smaller simulation which has similar dynamical properties and merger history. In the first instance, one could simply match halos at the final redshift output common to both simulations based upon their mass and properties at that epoch. To match the merger histories of the halos one would need to identify a method for matching entire merger trees of halos or use an indicator of the merger history, such as the 'formation time' when some fraction of the final mass is in place in a single object.

- Use the lightcone predictions to examine the redshift evolution of statistics of the galaxy population. An example is the mass-to-light ratios of halos, which has a characteristic dependency on halo mass that is a key prediction of the ΛCDM model. Using predictions from a lightcone mock catalogue one could examine the redshift evolution of this property. Additionally, one could examine how sensitive the mass-to-light ratio is to galaxy formation physics. For instance, one would expect the mass-to-light ratio to be sensitive to feedback processes from supernovae and active galactic nuclei. Predictions from the lightcone could be used to examine the extent to which such mechanisms affect the mass-to-light ratio. By constructing lightcones from simulations based upon alternative cosmologies (e.g. Li et al., 2012), one could then examine whether one would be able to use the mass-to-light ratio to distinguish between different cosmologies. These predictions could then be used to help interpret estimates of the mass-to-light ratio based upon measurements of groups of galaxies identified in high redshift galaxy surveys.

- One could investigate the possibility of constructing lightcone mock catalogues from hydrodynamical simulations. The interpolation techniques discussed in Chapter 3 of this thesis could be used to interpolate the position and velocities of the "galaxies" identified in a SPH simulation. However, just like lightcones based upon semi-analytic models, the complex time evolution of the intrinsic galaxy properties could make them difficult to interpolate. Estimates for properties such as the stellar mass and cold gas content of a galaxy could be made by interpolating the position of the star and gas particles and then using a technique, such as friends-of-friends, to determine which of these particles are associated with the galaxy.

- Exploit the multi-wavelength capability of the `GALFORM` model to construct suites of mock catalogues for galaxy surveys such as the GAMA survey, which will provide optical, far-infrared and HI gas content information for each galaxy.

Bibliography

Abazajian K. N., Adelman-McCarthy J. K., Agüeros M. A., Allam S. S., Allende Prieto C., An D., Anderson K. S. J., Anderson S. F., Annis J., Bahcall N. A., et al., 2009, ApJS, 182, 543

Abdalla F., Annis J., Bacon D., Bridle S., Castander F., Colless M., DePoy D., Diehl H. T., Eriksen M., Flaugher B., Frieman J., Gaztanaga E., Hogan C., Jouvel S., Kent S., Kirk D., Kron R., Kuhlmann S., Lahav O., Lawrence J., Lin H., Marriner J., Marshall J., Mohr J., Nichol R. C., Sako M., Saunders W., Soares-Santos M., Thomas D., Wechsler R., West A., Wu H., 2012, arXiv:astro-ph/1209.2451

Abdalla F. B., Banerji M., Lahav O., Rashkov V., 2011, MNRAS, 417, 1891

Adelberger K. L., Steidel C. C., Shapley A. E., Pettini M., 2003, ApJ, 584, 45

Agertz O., Moore B., Stadel J., Potter D., Miniati F., Read J., Mayer L., Gawryszczak A., Kravtsov A., Nordlund Å., Pearce F., Quilis V., Rudd D., Springel V., Stone J., Tasker E., Teyssier R., Wadsley J., Walder R., 2007, MNRAS, 380, 963

Alcock C., Allsman R. A., Alves D. R., Axelrod T. S., Becker A. C., Bennett D. P., Cook K. H., Dalal N., Drake A. J., Freeman K. C., Geha M., Griest K., Lehner M. J., Marshall S. L., Minniti D., Nelson C. A., Peterson B. A., Popowski P., Pratt M. R., Quinn P. J., Stubbs C. W., Sutherland W., Tomaney A. B., Vandehei T., Welch D., 2000, ApJ, 542, 281

Almeida C., Baugh C. M., Wake D. A., Lacey C. G., Benson A. J., Bower R. G., Pimbblet K., 2008, MNRAS, 386, 2145

Alpaslan M., Robotham A. S. G., Driver S., Norberg P., Peacock J. A., Baldry I., Bland-Hawthorn J., Brough S., Hopkins A. M., Kelvin L. S., Liske J., Loveday J., Merson A., Nichol R. C., Pimbblet K., 2012, MNRAS, 3012

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, arXiv:astro-ph/1203.3216

Arnouts S., Moscardini L., Vanzella E., Colombi S., Cristiani S., Fontana A., Giallongo E., Matarrese S., Saracco P., 2002, MNRAS, 329, 355

Arnouts S., Vandame B., Benoist C., Groenewegen M. A. T., da Costa L., Schirmer M., Mignani R. P., Slijkhuis R., Hatziminaoglou E., Hook R., Madejsky R., Rité C., Wicenec A., 2001, A&A, 379, 740

Assef R. J., Kochanek C. S., Brodwin M., Brown M. J. I., Caldwell N., Cool R. J., Eisenhardt P., Eisenstein D., Gonzalez A. H., Jannuzi B. T., Jones C., McKenzie E., Murray S. S., Stern D., 2008, ApJ, 676, 286

Babbedge T. S. R., Rowan-Robinson M., Gonzalez-Solares E., Polletta M., Berta S., Pérez-Fournon I., Oliver S., Salaman D. M., Irwin M., Weatherley S. J., 2004, MNRAS, 353, 654

Bahcall J. N., Tremaine S., 1981, ApJ, 244, 805

Bahcall N. A., Fan X., Cen R., 1997, ApJ, 485, L53

Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681

Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945

Balogh M. L., Christlein D., Zabludoff A. I., Zaritsky D., 2001, ApJ, 557, 117

Barnes J., Hut P., 1986, Nature, 324, 446

Baugh C. M., 2006, Reports on Progress in Physics, 69, 3101

—, 2008, Royal Society of London Philosophical Transactions Series A, 366, 4381

Baugh C. M., Gaztanaga E., Efstathiou G., 1995, MNRAS, 274, 1049

Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191

Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32

Bell E. F., Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 2003, MNRAS, 343, 367

Bell E. F., Naab T., McIntosh D. H., Somerville R. S., Caldwell J. A. R., Barden M., Wolf C., Rix H.-W., Beckwith S. V., Borch A., Häussler B., Heymans C., Jahnke K., Jogee S., Koposov S., Meisenheimer K., Peng C. Y., Sanchez S. F., Wisotzki L., 2006, ApJ, 640, 241

Bell E. F., Wolf C., Meisenheimer K., Rix H.-W., Borch A., Dye S., Kleinheinrich M., Wisotzki L., McIntosh D. H., 2004, ApJ, 608, 752

Benítez N., 2000, ApJ, 536, 571

Benítez N., Gaztañaga E., Miquel R., Castander F., Moles M., Crocce M., Fernández-Soto A., Fosalba P., Ballesteros F., Campa J., Cardiel-Sas L., Castilla J., Cristóbal-Hornillos D., Delfino M., Fernández E., Fernández-Sopuerta C., García-Bellido J., Lobo J. A., Martínez V. J., Ortiz A., Pacheco A., Paredes S., Pons-Bordería M. J., Sánchez E., Sánchez S. F., Varela J., de Vicente J. F., 2009, ApJ, 691, 241

Bennett C. L., Banday A. J., Gorski K. M., Hinshaw G., Jackson P., Keegstra P., Kogut A., Smoot G. F., Wilkinson D. T., Wright E. L., 1996, ApJ, 464, L1

Benson A. J., 2010, Phys. Rep., 495, 33

Benson A. J., Bower R., 2010, MNRAS, 405, 1573

—, 2011, MNRAS, 410, 2653

Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38

Benson A. J., Bower R. G., Frenk C. S., White S. D. M., 2000a, MNRAS, 314, 557

Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000b, MNRAS, 311, 793

Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2002, MNRAS, 333, 156

Benson A. J., Pearce F. R., Frenk C. S., Baugh C. M., Jenkins A., 2001, MNRAS, 320, 261

Berlind A. A., Frieman J., Weinberg D. H., Blanton M. R., Warren M. S., Abazajian K., Scranton R., Hogg D. W., Scoccimarro R., Bahcall N. A., Brinkmann J., Gott III J. R., Kleinman S. J., Krzesinski J., Lee B. C., Miller C. J., Nitta A., Schneider D. P., Tucker D. L., Zehavi I., SDSS Collaboration, 2006, ApJS, 167, 1

Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

Bernardi M., Sheth R. K., Annis J., Burles S., Eisenstein D. J., Finkbeiner D. P., Hogg D. W., Lupton R. H., Schlegel D. J., SubbaRao M., Bahcall N. A., Blakeslee J. P., Brinkmann J., Castander F. J., Connolly A. J., Csabai I., Doi M., Fukugita M., Frieman J., Heckman T., Hennessy G. S., Ivezić Ž., Knapp G. R., Lamb D. Q., McKay T., Munn J. A., Nichol R., Okamura S., Schneider D. P., Thakar A. R., York D. G., 2003, AJ, 125, 1866

Bett P., Eke V., Frenk C. S., Jenkins A., Helly J., Navarro J., 2007, MNRAS, 376, 215

Bielby R., Hudelot P., McCracken H. J., Ilbert O., Daddi E., Le Fèvre O., Gonzalez-Perez V., Kneib J.-P., Marmo C., Mellier Y., Salvato M., Sanders D. B., Willott C. J., 2012, A&A, 545, A23

Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846

Binney J., 1977, ApJ, 215, 483

—, 2004, MNRAS, 347, 1093

Binney J., Merrifield M., 1998, Galactic Astronomy

Binney J., Tremaine S., 1987, Galactic dynamics

Birnboim Y., Dekel A., 2003, MNRAS, 345, 349

Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, MNRAS, 360, 159

Blake C., Glazebrook K., 2003, ApJ, 594, 665

Blanc G. A., Lira P., Barrientos L. F., Aguirre P., Francke H., Taylor E. N., Quadri R., Marchesini D., Infante L., Gawiser E., Hall P. B., Willis J. P., Herrera D., Maza J., MUSYC Collaboration, 2008, ApJ, 681, 1099

Blanton M. R., Dalcanton J., Eisenstein D., Loveday J., Strauss M. A., SubbaRao M., Weinberg D. H., Anderson Jr. J. E., Annis J., Bahcall N. A., Bernardi M., Brinkmann J., Brunner R. J., Burles S., Carey L., Castander F. J., Connolly A. J., Csabai I., Doi M., Finkbeiner D., Friedman S., Frieman J. A., Fukugita M., Gunn J. E., Hennessy

G. S., Hindsley R. B., Hogg D. W., Ichikawa T., Ivezić Ž., Kent S., Knapp G. R., Lamb D. Q., Leger R. F., Long D. C., Lupton R. H., McKay T. A., Meiksin A., Merelli A., Munn J. A., Narayanan V., Newcomb M., Nichol R. C., Okamura S., Owen R., Pier J. R., Pope A., Postman M., Quinn T., Rockosi C. M., Schlegel D. J., Schneider D. P., Shimasaku K., Siegmund W. A., Smee S., Snir Y., Stoughton C., Stubbs C., Szalay A. S., Szokoly G. P., Thakar A. R., Tremonti C., Tucker D. L., Uomoto A., Vanden Berk D., Vogeley M. S., Waddell P., Yanny B., Yasuda N., York D. G., 2001, AJ, 121, 2358

Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005a, ApJ, 629, 143

Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J., 2005b, ApJ, 631, 208

Blanton M. R., Roweis S., 2007, AJ, 133, 734

Blitz L., Rosolowsky E., 2006, ApJ, 650, 933

Bolzonella M., Miralles J.-M., Pelló R., 2000, A&A, 363, 476

Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440

Borch A., Meisenheimer K., Bell E. F., Rix H.-W., Wolf C., Dye S., Kleinheinrich M., Kovacs Z., Wisotzki L., 2006, A&A, 453, 869

Bouchard A., Da Costa G. S., Jerjen H., 2009, AJ, 137, 3038

Bower R. G., 1991, MNRAS, 248, 332

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Brammer G. B., Whitaker K. E., van Dokkum P. G., Marchesini D., Labbé I., Franx M., Kriek M., Quadri R. F., Illingworth G., Lee K.-S., Muzzin A., Rudnick G., 2009, ApJ, 706, L173

Bressan A., Silva L., Granato G. L., 2002, A&A, 392, 377

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Bruzual A. G., Charlot S., 1993, ApJ, 405, 538

Budavári T., 2009, ApJ, 695, 747

Bundy K., Fukugita M., Ellis R. S., Targett T. A., Belli S., Kodama T., 2009, ApJ, 697, 1369

Cabré A., Gaztañaga E., 2009a, MNRAS, 393, 1183

—, 2009b, MNRAS, 396, 1119

Cai Y.-C., Angulo R. E., Baugh C. M., Cole S., Frenk C. S., Jenkins A., 2009, MNRAS, 395, 1185

Capak P., Cowie L. L., Hu E. M., Barger A. J., Dickinson M., Fernandez E., Giavalisco M., Komiyama Y., Kretchmer C., McNally C., Miyazaki S., Okamura S., Stern D., 2004, AJ, 127, 180

Capozzi D., Collins C. A., Stott J. P., Hilton M., 2012, MNRAS, 419, 2821

Carlberg R. G., Yee H. K. C., Ellingson E., Abraham R., Gravel P., Morris S., Pritchet C. J., 1996, ApJ, 462, 32

Carlson J., White M., 2010, ApJS, 190, 311

Cavaliere A., Fusco-Femiano R., 1976, A&A, 49, 137

Chabrier G., 2003, PASP, 115, 763

Coil A. L., Davis M., Madgwick D. S., Newman J. A., Conselice C. J., Cooper M., Ellis R. S., Faber S. M., Finkbeiner D. P., Guhathakurta P., Kaiser N., Koo D. C., Phillips A. C., Steidel C. C., Weiner B. J., Willmer C. N. A., Yan R., 2004a, ApJ, 609, 525

Coil A. L., Newman J. A., Kaiser N., Davis M., Ma C.-P., Kocevski D. D., Koo D. C., 2004b, ApJ, 617, 765

Cole S., 1991, ApJ, 367, 45

Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781

Cole S., Hatton S., Weinberg D. H., Frenk C. S., 1998, MNRAS, 300, 945

Cole S., Lacey C., 1996, MNRAS, 281, 716

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Cole S., Norberg P., Baugh C. M., Frenk C. S., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Cross N., Dalton G., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Glazebrook K., Jackson C., Lahav O., Lewis I., Lumsden S., Maddox S., Madgwick D., Peacock J. A., Peterson B. A., Sutherland W., Taylor K., 2001, MNRAS, 326, 255

Cole S., Percival W. J., Peacock J. A., Norberg P., Baugh C. M., Frenk C. S., Baldry I., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Cross N. J. G., Dalton G., Eke V. R., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Glazebrook K., Jackson C., Jenkins A., Lahav O., Lewis I., Lumsden S., Maddox S., Madgwick D., Peterson B. A., Sutherland W., Taylor K., 2005, MNRAS, 362, 505

Colless M., Dalton G., Maddox S., Sutherland W., Norberg P., Cole S., Bland-Hawthorn J., Bridges T., Cannon R., Collins C., Couch W., Cross N., Deeley K., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Frenk C. S., Glazebrook K., Jackson C., Lahav O., Lewis I., Lumsden S., Madgwick D., Peacock J. A., Peterson B. A., Price I., Seaborne M., Taylor K., 2001, MNRAS, 328, 1039

Colless M., Peterson B. A., Jackson C., Peacock J. A., Cole S., Norberg P., Baldry I. K., Baugh C. M., Bland-Hawthorn J., Bridges T., Cannon R., Collins C., Couch W., Cross N., Dalton G., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Frenk C. S., Glazebrook K., Lahav O., Lewis I., Lumsden S., Maddox S., Madgwick D., Sutherland W., Taylor K., 2003, ArXiv Astrophysics e-prints

Collister A., Lahav O., Blake C., Cannon R., Croom S., Drinkwater M., Edge A., Eisenstein D., Loveday J., Nichol R., Pimbblet K., de Propris R., Roseboom I., Ross N., Schneider D. P., Shanks T., Wake D., 2007, MNRAS, 375, 68

Collister A. A., Lahav O., 2004, PASP, 116, 345

Connolly A. J., Csabai I., Szalay A. S., Koo D. C., Kron R. G., Munn J. A., 1995, AJ, 110, 2655

Cooper M. C., Newman J. A., Coil A. L., Croton D. J., Gerke B. F., Yan R., Davis M., Faber S. M., Guhathakurta P., Koo D. C., Weiner B. J., Willmer C. N. A., 2007, MNRAS, 376, 1445

Cooper M. C., Newman J. A., Weiner B. J., Yan R., Willmer C. N. A., Bundy K., Coil A. L., Conselice C. J., Davis M., Faber S. M., Gerke B. F., Guhathakurta P., Koo D. C., Noeske K. G., 2008, MNRAS, 383, 1058

Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, AJ, 112, 839

Crook A. C., Huchra J. P., Martimbeau N., Masters K. L., Jarrett T., Macri L. M., 2007, ApJ, 655, 790

Croton D. J., Farrar G. R., Norberg P., Colless M., Peacock J. A., Baldry I. K., Baugh C. M., Bland-Hawthorn J., Bridges T., Cannon R., Cole S., Collins C., Couch W., Dalton G., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Frenk C. S., Glazebrook K., Jackson C., Lahav O., Lewis I., Lumsden S., Maddox S., Madgwick D., Peterson B. A., Sutherland W., Taylor K., 2005, MNRAS, 356, 1155

Croton D. J., Springel V., White S. D. M., De Lucia G., Frenk C. S., Gao L., Jenkins A., Kauffmann G., Navarro J. F., Yoshida N., 2006, MNRAS, 365, 11

Csabai I., Budavári T., Connolly A. J., Szalay A. S., Győry Z., Benítez N., Annis J., Brinkmann J., Eisenstein D., Fukugita M., Gunn J., Kent S., Lupton R., Nichol R. C., Stoughton C., 2003, AJ, 125, 580

Daddi E., Cimatti A., Renzini A., Fontana A., Mignoli M., Pozzetti L., Tozzi P., Zamorani G., 2004a, ApJ, 617, 746

Daddi E., Cimatti A., Renzini A., Vernet J., Conselice C., Pozzetti L., Mignoli M., Tozzi P., Broadhurst T., di Serego Alighieri S., Fontana A., Nonino M., Rosati P., Zamorani G., 2004b, ApJ, 600, L127

Daddi E., Dickinson M., Chary R., Pope A., Morrison G., Alexander D. M., Bauer F. E., Brandt W. N., Giavalisco M., Ferguson H., Lee K.-S., Lehmer B. D., Papovich C., Renzini A., 2005a, ApJ, 631, L13

Daddi E., Renzini A., Pirzkal N., Cimatti A., Malhotra S., Stiavelli M., Xu C., Pasquali A., Rhoads J. E., Brusa M., di Serego Alighieri S., Ferguson H. C., Koekemoer A. M., Moustakas L. A., Panagia N., Windhorst R. A., 2005b, ApJ, 626, 680

Danese L., de Zotti G., di Tullio G., 1980, A&A, 82, 322

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

Davis M., Faber S. M., Newman J., Phillips A. C., Ellis R. S., Steidel C. C., Conselice C., Coil A. L., Finkbeiner D. P., Koo D. C., Guhathakurta P., Weiner B., Schiavon R., Willmer C., Kaiser N., Luppino G. A., Wirth G., Connolly A., Eisenhardt P., Cooper M., Gerke B., 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4834, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Guhathakurta P., ed., pp. 161–172

Davis M., Peebles P. J. E., 1983, ApJ, 267, 465

de Bernardis P., Ade P. A. R., Bock J. J., Bond J. R., Borrill J., Boscaleri A., Coble K., Crill B. P., De Gasperis G., Farese P. C., Ferreira P. G., Ganga K., Giacometti M., Hivon E., Hristov V. V., Iacoangeli A., Jaffe A. H., Lange A. E., Martinis L., Masi S., Mason P. V., Mauskopf P. D., Melchiorri A., Miglio L., Montroy T., Netterfield C. B., Pascale E., Piacentini F., Pogosyan D., Prunet S., Rao S., Romeo G., Ruhl J. E., Scaramuzzi F., Sforna D., Vittorio N., 2000, Nature, 404, 955

De Propris R., Colless M., Driver S. P., Couch W., Peacock J. A., Baldry I. K., Baugh C. M., Bland-Hawthorn J., Bridges T., Cannon R., Cole S., Collins C., Cross N., Dalton G. B., Efstathiou G., Ellis R. S., Frenk C. S., Glazebrook K., Hawkins E., Jackson C., Lahav O., Lewis I., Lumsden S., Maddox S., Madgwick D. S., Norberg P., Percival W., Peterson B., Sutherland W., Taylor K., 2003, MNRAS, 342, 725

De Propris R., Conselice C. J., Liske J., Driver S. P., Patton D. R., Graham A. W., Allen P. D., 2007, ApJ, 666, 212

de Ravel L., Kampczyk P., Le Fèvre O., Lilly S. J., Tasca L., Tresse L., Lopez-Sanjuan C., Bolzonella M., Kovac K., Abbas U., Bardelli S., Bongiorno A., Caputi K., Contini T., Coppa G., Cucciati O., de la Torre S., Dunlop J. S., Franzetti P., Garilli B., Iovino A., Kneib J. ., Koekemoer A. M., Knobel C., Lamareille F., Le Borgne J. ., Le Brun V., Leauthaud A., Maier C., Mainieri V., Mignoli M., Pello R., Peng Y., Perez Montero E., Ricciardelli E., Scodeggio M., Silverman J. D., Tanaka M., Vergani D., Zamorani

G., Zucca E., Bottini D., Cappi A., Carollo C. M., Cassata P., Cimatti A., Fumana M., Guzzo L., Maccagni D., Marinoni C., McCracken H. J., Memeo P., Meneux B., Oesch P., Porciani C., Pozzetti L., Renzini A., Scaramella R., Scarlata C., 2011, arXiv:astro-ph/1104.5470

de Ravel L., Le Fèvre O., Tresse L., Bottini D., Garilli B., Le Brun V., Maccagni D., Scaramella R., Scodeggio M., Vettolani G., Zanichelli A., Adami C., Arnouts S., Bardelli S., Bolzonella M., Cappi A., Charlot S., Ciliegi P., Contini T., Foucaud S., Franzetti P., Gavignaud I., Guzzo L., Ilbert O., Iovino A., Lamareille F., McCracken H. J., Marano B., Marinoni C., Mazure A., Meneux B., Merighi R., Paltani S., Pellò R., Pollo A., Pozzetti L., Radovich M., Vergani D., Zamorani G., Zucca E., Bondi M., Bongiorno A., Brinchmann J., Cucciati O., de La Torre S., Gregorini L., Memeo P., Perez-Montero E., Mellier Y., Merluzzi P., Temporin S., 2009, A&A, 498, 379

Dekel A., Silk J., 1986, ApJ, 303, 39

Deng X.-F., Chen Y.-Q., Jiang P., 2011, MNRAS, 417, 453

Diaferio A., Kauffmann G., Colberg J. M., White S. D. M., 1999, MNRAS, 307, 537

Dickinson M., Papovich C., Ferguson H. C., Budavári T., 2003, ApJ, 587, 25

Dotter A., Chaboyer B., Ferguson J. W., Lee H.-c., Worthey G., Jevremović D., Baron E., 2007, ApJ, 666, 403

Dressler A., 1980, ApJ, 236, 351

Drinkwater M. J., Jurek R. J., Blake C., Woods D., Pimbblet K. A., Glazebrook K., Sharp R., Pracy M. B., Brough S., Colless M., Couch W. J., Croom S. M., Davis T. M., Forbes D., Forster K., Gilbank D. G., Gladders M., Jelliffe B., Jones N., Li I.-H., Madore B., Martin D. C., Poole G. B., Small T., Wisnioski E., Wyder T., Yee H. K. C., 2010, MNRAS, 401, 1429

Driver S. P., Hill D. T., Kelvin L. S., Robotham A. S. G., Liske J., Norberg P., Baldry I. K., Bamford S. P., Hopkins A. M., Loveday J., Peacock J. A., Andrae E., Bland-Hawthorn J., Brough S., Brown M. J. I., Cameron E., Ching J. H. Y., Colless M., Conselice C. J., Croom S. M., Cross N. J. G., de Propris R., Dye S., Drinkwater M. J., Ellis S., Graham A. W., Grootes M. W., Gunawardhana M., Jones D. H., van Kampen E., Maraston C., Nichol R. C., Parkinson H. R., Phillipps S., Pimbblet K., Popescu

C. C., Prescott M., Roseboom I. G., Sadler E. M., Sansom A. E., Sharp R. G., Smith D. J. B., Taylor E., Thomas D., Tuffs R. J., Wijesinghe D., Dunne L., Frenk C. S., Jarvis M. J., Madore B. F., Meyer M. J., Seibert M., Staveley-Smith L., Sutherland W. J., Warren S. J., 2011, MNRAS, 413, 971

Driver S. P., Norberg P., Baldry I. K., Bamford S. P., Hopkins A. M., Liske J., Loveday J., Peacock J. A., Hill D. T., Kelvin L. S., Robotham A. S. G., Cross N. J. G., Parkinson H. R., Prescott M., Conselice C. J., Dunne L., Brough S., Jones H., Sharp R. G., van Kampen E., Oliver S., Roseboom I. G., Bland-Hawthorn J., Croom S. M., Ellis S., Cameron E., Cole S., Frenk C. S., Couch W. J., Graham A. W., Proctor R., De Propris R., Doyle I. F., Edmondson E. M., Nichol R. C., Thomas D., Eales S. A., Jarvis M. J., Kuijken K., Lahav O., Madore B. F., Seibert M., Meyer M. J., Staveley-Smith L., Phillipps S., Popescu C. C., Sansom A. E., Sutherland W. J., Tuffs R. J., Warren S. J., 2009, Astronomy and Geophysics, 50, 050000

Dubinski J., Carlberg R. G., 1991, ApJ, 378, 496

Dunkley J., Komatsu E., Nolta M. R., Spergel D. N., Larson D., Hinshaw G., Page L., Bennett C. L., Gold B., Jarosik N., Weiland J. L., Halpern M., Hill R. S., Kogut A., Limon M., Meyer S. S., Tucker G. S., Wollack E., Wright E. L., 2009, ApJS, 180, 306

Efstathiou G., Lake G., Negroponte J., 1982, MNRAS, 199, 1069

Efstathiou G., Sutherland W. J., Maddox S. J., 1990, Nature, 348, 705

Einasto J., 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87

Eisenstein D. J., Hu W., 1998, ApJ, 496, 605

Eisenstein D. J., Weinberg D. H., Agol E., Aihara H., Allende Prieto C., Anderson S. F., Arns J. A., Aubourg É., Bailey S., Balbinot E., et al., 2011, AJ, 142, 72

Eisenstein D. J., Zehavi I., Hogg D. W., Scoccimarro R., Blanton M. R., Nichol R. C., Scranton R., Seo H.-J., Tegmark M., Zheng Z., Anderson S. F., Annis J., Bahcall N., Brinkmann J., Burles S., Castander F. J., Connolly A., Csabai I., Doi M., Fukugita M., Frieman J. A., Glazebrook K., Gunn J. E., Hendry J. S., Hennessy G., Ivezić Z., Kent S., Knapp G. R., Lin H., Loh Y.-S., Lupton R. H., Margon B., McKay T. A., Meiksin A., Munn J. A., Pope A., Richmond M. W., Schlegel D., Schneider D. P., Shimasaku K., Stoughton C., Strauss M. A., SubbaRao M., Szalay A. S., Szapudi I., Tucker D. L., Yanny B., York D. G., 2005, ApJ, 633, 560

Eke V. R., Baugh C. M., Cole S., Frenk C. S., Norberg P., Peacock J. A., Baldry I. K., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Dalton G., de Propris R., Driver S. P., Efstathiou G., Ellis R. S., Glazebrook K., Jackson C., Lahav O., Lewis I., Lumsden S., Maddox S., Madgwick D., Peterson B. A., Sutherland W., Taylor K., 2004a, MNRAS, 348, 866

Eke V. R., Cole S., Frenk C. S., 1996, MNRAS, 282, 263

Eke V. R., Cole S., Frenk C. S., Patrick Henry J., 1998a, MNRAS, 298, 1145

Eke V. R., Frenk C. S., Baugh C. M., Cole S., Norberg P., Peacock J. A., Baldry I. K., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Dalton G., de Propris R., Driver S. P., Efstathiou G., Ellis R. S., Glazebrook K., Jackson C. A., Lahav O., Lewis I., Lumsden S., Maddox S. J., Madgwick D., Peterson B. A., Sutherland W., Taylor K., 2004b, MNRAS, 355, 769

Eke V. R., Navarro J. F., Frenk C. S., 1998b, ApJ, 503, 569

Elbaz D., Daddi E., Le Borgne D., Dickinson M., Alexander D. M., Chary R.-R., Starck J.-L., Brandt W. N., Kitzbichler M., MacDonald E., Nonino M., Popesso P., Stern D., Vanzella E., 2007, A&A, 468, 33

Ellison S. L., Patton D. R., Simard L., McConnachie A. W., 2008, ApJ, 672, L107

Elston R., Rieke G. H., Rieke M. J., 1988, ApJ, 331, L77

Erb D. K., Shapley A. E., Pettini M., Steidel C. C., Reddy N. A., Adelberger K. L., 2006, ApJ, 644, 813

Evrard A. E., Henry J. P., 1991, ApJ, 383, 95

Faber S. M., Willmer C. N. A., Wolf C., Koo D. C., Weiner B. J., Newman J. A., Im M., Coil A. L., Conroy C., Cooper M. C., Davis M., Finkbeiner D. P., Gerke B. F., Gebhardt K., Groth E. J., Guhathakurta P., Harker J., Kaiser N., Kassin S., Kleinheinrich M., Konidaris N. P., Kron R. G., Lin L., Luppino G., Madgwick D. S., Meisenheimer K., Noeske K. G., Phillips A. C., Sarajedini V. L., Schiavon R. P., Simard L., Szalay A. S., Vogt N. P., Yan R., 2007, ApJ, 665, 265

Fanidakis N., Baugh C. M., Benson A. J., Bower R. G., Cole S., Done C., Frenk C. S., 2011, MNRAS, 410, 53

Fardal M. A., Katz N., Gardner J. P., Hernquist L., Weinberg D. H., Davé R., 2001, ApJ, 562, 605

Fardal M. A., Katz N., Weinberg D. H., Davé R., 2007, MNRAS, 379, 985

Feldmann R., Carollo C. M., Porciani C., Lilly S. J., Capak P., Taniguchi Y., Le Fèvre O., Renzini A., Scoville N., Ajiki M., Aussel H., Contini T., McCracken H., Mobasher B., Murayama T., Sanders D., Sasaki S., Scarlata C., Scodeggio M., Shioya Y., Silverman J., Takahashi M., Thompson D., Zamorani G., 2006, MNRAS, 372, 565

Ferguson H. C., Dickinson M., Williams R., 2000, ARA&A, 38, 667

Ferrara A., Bianchi S., Cimatti A., Giovanardi C., 1999, ApJS, 123, 437

Figer D. F., Kim S. S., Morris M., Serabyn E., Rich R. M., McLean I. S., 1999, ApJ, 525, 750

Fioc M., Rocca-Volmerange B., 1999, ArXiv Astrophysics e-prints

Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195

Font A. S., Bower R. G., McCarthy I. G., Benson A. J., Frenk C. S., Helly J. C., Lacey C. G., Baugh C. M., Cole S., 2008, MNRAS, 389, 1619

Fontana A., Pozzetti L., Donnarumma I., Renzini A., Cimatti A., Zamorani G., Menci N., Daddi E., Giallongo E., Mignoli M., Perna C., Salimbeni S., Saracco P., Broadhurst T., Cristiani S., D'Odorico S., Gilmozzi R., 2004, A&A, 424, 23

Forman W., Jones C., Churazov E., Markevitch M., Nulsen P., Vikhlinin A., Begelman M., Böhringer H., Eilek J., Heinz S., Kraft R., Owen F., Pahre M., 2007, ApJ, 665, 1057

Foster C., Hopkins A. M., Gunawardhana M., Lara-Lopez M. A., Sharp R. G., Steele O., Taylor E. N., Driver S. P., Baldryi I. K., Bamford S. P., Liske J., Loveday J., Norberg P., Peacock J. A., Alpaslan M., Bauer A. E., Bland-Hawthorn J., Brough S., Cameron E., Colless M., Conselice C. J., Croom S. M., Frenk C. S., Hill D. T., Jones D. H., Kelvin L. S., Kuijken K., Nichol R. C., Owers M. S., Parkinson H. R., Pimbblet K. A., Popescu C. C., Prescott M., Robotham A. S. G., Lopez-Sanchez A. R., Sutherland W. J., Thomas D., Tuffs R. J., van Kampen E., Wijesinghe D., 2012, arXiv:astro-ph/1209.1636

Franx M., Labbé I., Rudnick G., van Dokkum P. G., Daddi E., Förster Schreiber N. M., Moorwood A., Rix H.-W., Röttgering H., van der Wel A., van der Werf P., van Starkenburg L., 2003, ApJ, 587, L79

Frenk C. S., White S. D. M., Bode P., Bond J. R., Bryan G. L., Cen R., Couchman H. M. P., Evrard A. E., Gnedin N., Jenkins A., Khokhlov A. M., Klypin A., Navarro J. F., Norman M. L., Ostriker J. P., Owen J. M., Pearce F. R., Pen U.-L., Steinmetz M., Thomas P. A., Villumsen J. V., Wadsley J. W., Warren M. S., Xu G., Yepes G., 1999, ApJ, 525, 554

Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41

Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66

Gerke B. F., Newman J. A., Davis M., Coil A. L., Cooper M. C., Dutton A. A., Faber S. M., Guhathakurta P., Konidaris N., Koo D. C., Lin L., Noeske K., Phillips A. C., Rosario D. J., Weiner B. J., Willmer C. N. A., Yan R., 2012, ApJ, 751, 50

Gerke B. F., Newman J. A., Davis M., Marinoni C., Yan R., Coil A. L., Conroy C., Cooper M. C., Faber S. M., Finkbeiner D. P., Guhathakurta P., Kaiser N., Koo D. C., Phillips A. C., Weiner B. J., Willmer C. N. A., 2005, ApJ, 625, 6

Giuricin G., Marinoni C., Ceriani L., Pisani A., 2000, ApJ, 543, 178

González J. E., Lacey C. G., Baugh C. M., Frenk C. S., Benson A. J., 2009, MNRAS, 397, 1254

Gonzalez-Perez V., Baugh C. M., Lacey C. G., Almeida C., 2009, MNRAS, 398, 497

Gorski K. M., Hinshaw G., Banday A. J., Bennett C. L., Wright E. L., Kogut A., Smoot G. F., Lubin P., 1994, ApJ, 430, L89

Goto T., Yamauchi C., Fujita Y., Okamura S., Sekiguchi M., Smail I., Bernardi M., Gomez P. L., 2003, MNRAS, 346, 601

Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, ApJ, 542, 710

Grazian A., Salimbeni S., Pentericci L., Fontana A., Nonino M., Vanzella E., Cristiani S., de Santis C., Gallozzi S., Giallongo E., Santini P., 2007, A&A, 465, 393

Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 385

Grützbauch R., Conselice C. J., Bauer A. E., Bluck A. F. L., Chuter R. W., Buitrago F., Mortlock A., Weinzirl T., Jogee S., 2011, MNRAS, 418, 938

Guth A. H., 1981, Phys. Rev. D, 23, 347

Halverson N. W., Leitch E. M., Pryke C., Kovac J., Carlstrom J. E., Holzapfel W. L., Dragovan M., Cartwright J. K., Mason B. S., Padin S., Pearson T. J., Readhead A. C. S., Shepherd M. C., 2002, ApJ, 568, 38

Hambly N. C., Irwin M. J., MacGillivray H. T., 2001, MNRAS, 326, 1295

Hamilton A. J. S., 1993, ApJ, 417, 19

Hartley W. G., Lane K. P., Almaini O., Cirasuolo M., Foucaud S., Simpson C., Maddox S., Smail I., Conselice C. J., McLure R. J., Dunlop J. S., 2008, MNRAS, 391, 1301

Hawkins E., Maddox S., Cole S., Lahav O., Madgwick D. S., Norberg P., Peacock J. A., Baldry I. K., Baugh C. M., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Dalton G., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Frenk C. S., Glazebrook K., Jackson C., Jones B., Lewis I., Lumsden S., Percival W., Peterson B. A., Sutherland W., Taylor K., 2003, MNRAS, 346, 78

Hayashi M., Motohara K., Shimasaku K., Onodera M., Uchimoto Y. K., Kashikawa N., Yoshida M., Okamura S., Ly C., Malkan M. A., 2009, ApJ, 691, 140

Hayashi M., Shimasaku K., Motohara K., Yoshida M., Okamura S., Kashikawa N., 2007, ApJ, 660, 72

Heckman T. M., Lehnert M. D., Strickland D. K., Armus L., 2000, ApJS, 129, 493

Heisler J., Tremaine S., Bahcall J. N., 1985, ApJ, 298, 8

Helly J. C., Cole S., Frenk C. S., Baugh C. M., Benson A., Lacey C., Pearce F. R., 2003, MNRAS, 338, 913

Henriques B. M. B., White S. D. M., Lemson G., Thomas P. A., Guo Q., Marleau G.-D., Overzier R. A., 2012, MNRAS, 421, 2904

Hernquist L., Mihos J. C., 1995, ApJ, 448, 41

Hopkins P. F., Bundy K., Croton D., Hernquist L., Keres D., Khochfar S., Stewart K., Wetzel A., Younger J. D., 2010, ApJ, 715, 202

Huchra J. P., Geller M. J., 1982, ApJ, 257, 423

Ideue Y., Nagao T., Taniguchi Y., Shioya Y., Saito T., Murayama T., Sasaki S., Trump J. R., Koekemoer A. M., Aussel H., Capak P., Ilbert O., McCracken H., Mobasher B., Salvato M., Sanders D. B., Scoville N., 2009, ApJ, 700, 971

Ideue Y., Taniguchi Y., Nagao T., Shioya Y., Kajisawa M., Trump J. R., Vergani D., Iovino A., Koekemoer A. M., Le Fèvre O., Ilbert O., Scoville N. Z., 2012, ApJ, 747, 42

Ilbert O., Salvato M., Le Floc'h E., Aussel H., Capak P., McCracken H. J., Mobasher B., Kartaltepe J., Scoville N., Sanders D. B., Arnouts S., Bundy K., Cassata P., Kneib J.-P., Koekemoer A., Le Fèvre O., Lilly S., Surace J., Taniguchi Y., Tasca L., Thompson D., Tresse L., Zamojski M., Zamorani G., Zucca E., 2010, ApJ, 709, 644

Imai K., Pearson C. P., Matsuhara H., Wada T., Oyabu S., Takagi T., Fujishiro N., Hanami H., 2008, ApJ, 683, 45

Iovino A., McCracken H. J., Garilli B., Foucaud S., Le Fèvre O., Maccagni D., Saracco P., Bardelli S., Busarello G., Scodeggio M., Zanichelli A., Paioro L., Bottini D., Le Brun V., Picat J. P., Scaramella R., Tresse L., Vettolani G., Adami C., Arnaboldi M., Arnouts S., Bolzonella M., Cappi A., Charlot S., Ciliegi P., Contini T., Franzetti P., Gavignaud I., Guzzo L., Ilbert O., Marano B., Marinoni C., Mazure A., Meneux B., Merighi R., Paltani S., Pellò R., Pollo A., Pozzetti L., Radovich M., Zamorani G., Zucca E., Bertin E., Bondi M., Bongiorno A., Cucciati O., Gregorini L., Mathez G., Mellier Y., Merluzzi P., Ripepi V., Rizzo D., 2005, A&A, 442, 423

Jarrett T. H., Chester T., Cutri R., Schneider S., Skrutskie M., Huchra J. P., 2000, AJ, 119, 2498

Jenkins A., Frenk C. S., Pearce F. R., Thomas P. A., Colberg J. M., White S. D. M., Couchman H. M. P., Peacock J. A., Efstathiou G., Nelson A. H., 1998, ApJ, 499, 20

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372

Jing Y. P., Mo H. J., Boerner G., 1998, ApJ, 494, 1

Jones D. H., Peterson B. A., Colless M., Saunders W., 2006, MNRAS, 369, 25

Jones D. H., Read M. A., Saunders W., Colless M., Jarrett T., Parker Q. A., Fairall A. P., Mauch T., Sadler E. M., Watson F. G., Burton D., Campbell L. A., Cass P., Croom S. M., Dawe J., Fiegert K., Frankcombe L., Hartley M., Huchra J., James D., Kirby E., Lahav O., Lucey J., Mamon G. A., Moore L., Peterson B. A., Prior S., Proust D., Russell K., Safouris V., Wakamatsu K.-I., Westra E., Williams M., 2009, MNRAS, 399, 683

Jones D. H., Saunders W., Colless M., Read M. A., Parker Q. A., Watson F. G., Campbell L. A., Burkey D., Mauch T., Moore L., Hartley M., Cass P., James D., Russell K., Fiegert K., Dawe J., Huchra J., Jarrett T., Lahav O., Lucey J., Mamon G. A., Proust D., Sadler E. M., Wakamatsu K.-i., 2004, MNRAS, 355, 747

Kaiser N., 1987, MNRAS, 227, 1

Kaiser N., Aussel H., Burke B. E., Boesgaard H., Chambers K., Chun M. R., Heasley J. N., Hodapp K.-W., Hunt B., Jedicke R., Jewitt D., Kudritzki R., Luppino G. A., Maberry M., Magnier E., Monet D. G., Onaka P. M., Pickles A. J., Rhoads P. H. H., Simon T., Szalay A., Szapudi I., Tholen D. J., Tonry J. L., Waterson M., Wick J., 2002, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4836, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Tyson J. A., Wolff S., eds., pp. 154–164

Kashikawa N., Shimasaku K., Yasuda N., Ajiki M., Akiyama M., Ando H., Aoki K., Doi M., Fujita S. S., Furusawa H., Hayashino T., Iwamuro F., Iye M., Karoji H., Kobayashi N., Kodaira K., Kodama T., Komiyama Y., Matsuda Y., Miyazaki S., Mizumoto Y., Morokuma T., Motohara K., Murayama T., Nagao T., Nariai K., Ohta K., Okamura S., Ouchi M., Sasaki T., Sato Y., Sekiguchi K., Shioya Y., Tamura H., Taniguchi Y., Umemura M., Yamada T., Yoshida M., 2004, PASJ, 56, 1011

Kauffmann G., Heckman T. M., White S. D. M., Charlot S., Tremonti C., Brinchmann J., Bruzual G., Peng E. W., Seibert M., Bernardi M., Blanton M., Brinkmann J., Castander F., Csábai I., Fukugita M., Ivezic Z., Munn J. A., Nichol R. C., Padmanabhan N., Thakar A. R., Weinberg D. H., York D., 2003a, MNRAS, 341, 33

Kauffmann G., Heckman T. M., White S. D. M., Charlot S., Tremonti C., Peng E. W.,

Seibert M., Brinkmann J., Nichol R. C., SubbaRao M., York D., 2003b, MNRAS, 341, 54

Kauffmann G., White S. D. M., Guiderdoni B., 1993, MNRAS, 264, 201

Kay S. T., Bower R. G., 1999, MNRAS, 308, 664

Keenan R. C., Trouille L., Barger A. J., Cowie L. L., Wang W.-H., 2010, ApJS, 186, 94

Kennicutt Jr. R. C., 1983, ApJ, 272, 54

—, 1989, ApJ, 344, 685

—, 1998, ApJ, 498, 541

Kennicutt Jr. R. C., Calzetti D., Walter F., Helou G., Hollenbach D. J., Armus L., Bendo G., Dale D. A., Draine B. T., Engelbracht C. W., Gordon K. D., Prescott M. K. M., Regan M. W., Thornley M. D., Bot C., Brinks E., de Blok E., de Mello D., Meyer M., Moustakas J., Murphy E. J., Sheth K., Smith J. D. T., 2007, ApJ, 671, 333

Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

King I. R., 1972, ApJ, 174, L123

Kitzbichler M. G., White S. D. M., 2007, MNRAS, 376, 2

Komatsu E., Smith K. M., Dunkley J., Bennett C. L., Gold B., Hinshaw G., Jarosik N., Larson D., Nolta M. R., Page L., Spergel D. N., Halpern M., Hill R. S., Kogut A., Limon M., Meyer S. S., Odegard N., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2011, ApJS, 192, 18

Kong X., Daddi E., Arimoto N., Renzini A., Broadhurst T., Cimatti A., Ikuta C., Ohta K., da Costa L., Olsen L. F., Onodera M., Tamura N., 2006, ApJ, 638, 72

Kormendy J., 2001, in Astronomical Society of the Pacific Conference Series, Vol. 230, Galaxy Disks and Disk Galaxies, Funes J. G., Corsini E. M., eds., pp. 247–256

Kriek M., van Dokkum P. G., Franx M., Förster Schreiber N. M., Gawiser E., Illingworth G. D., Labbé I., Marchesini D., Quadri R., Rix H.-W., Rudnick G., Toft S., van der Werf P., Wuyts S., 2006, ApJ, 645, 44

Kriek M., van Dokkum P. G., Whitaker K. E., Labbé I., Franx M., Brammer G. B., 2011, ApJ, 743, 168

Kroupa P., 2001, MNRAS, 322, 231

Krumholz M. R., McKee C. F., 2005, ApJ, 630, 250

Krumholz M. R., McKee C. F., Tumlinson J., 2009, ApJ, 699, 850

Lacey C., Cole S., 1993, MNRAS, 262, 627

Lacey C. G., Baugh C. M., Frenk C. S., Benson A. J., Orsi A., Silva L., Granato G. L., Bressan A., 2010, MNRAS, 405, 2

Lacey C. G., Baugh C. M., Frenk C. S., Silva L., Granato G. L., Bressan A., 2008, MNRAS, 385, 1155

Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011a, MNRAS, 418, 1649

Lagos C. d. P., Bayet E., Baugh C. M., Lacey C. G., Bell T., Fanidakis N., Geach J., 2012, arXiv:astro-ph/1204.0795

Lagos C. D. P., Lacey C. G., Baugh C. M., Bower R. G., Benson A. J., 2011b, MNRAS, 416, 1566

Lambas D. G., Alonso S., Mesa V., O'Mill A. L., 2012, A&A, 539, A45

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Lane K. P., Almaini O., Foucaud S., Simpson C., Smail I., McLure R. J., Conselice C. J., Cirasuolo M., Page M. J., Dunlop J. S., Hirst P., Watson M. G., Sekiguchi K., 2007, MNRAS, 379, L25

Larson R. B., 1998, MNRAS, 301, 569

—, 2005, MNRAS, 359, 211

Laureijs R., Amiaux J., Arduini S., Augueres J. ., Brinchmann J., Cole R., Cropper M., Dabin C., Duvet L., Ealet A., et al., 2011, arXiv:astro-ph/1110.3193

Le Fèvre O., Mellier Y., McCracken H. J., Foucaud S., Gwyn S., Radovich M., Dantel-Fort M., Bertin E., Moreau C., Cuillandre J.-C., Pierre M., Le Brun V., Mazure A., Tresse L., 2004, A&A, 417, 839

Le Fèvre O., Vettolani G., Maccagni D., Picat J.-P., Garilli B., Tresse L., Adami C., Arnaboldi M., Arnouts S., Bardelli S., Bolzonella M., Bottini D., Buzzarello G., Charlot S., Chincarini G., Contini T., Foucaud S., Franzetti P., Guzzo L., Gwyn S., Ilbert O., Iovino A., Le Brun V., Longhetti M., Marinoni C., Methez G., Mazure A., McCracken H., Mellier Y., Meneux B., Merluzzi P., Paltani S., Pellò R., Pollo A., Radovich M., Rippepi P., Rizzo D., Scaramella R., Scodeggio M., Zamorani G., Zanichelli A., Zucca E., 2003, The Messenger, 111, 18

Lee H.-c., Worthey G., Dotter A., Chaboyer B., Jevremović D., Baron E., Briley M. M., Ferguson J. W., Coelho P., Trager S. C., 2009, ApJ, 694, 902

Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782

Lewis I., Balogh M., De Propris R., Couch W., Bower R., Offer A., Bland-Hawthorn J., Baldry I. K., Baugh C., Bridges T., Cannon R., Cole S., Colless M., Collins C., Cross N., Dalton G., Driver S. P., Efstathiou G., Ellis R. S., Frenk C. S., Glazebrook K., Hawkins E., Jackson C., Lahav O., Lumsden S., Maddox S., Madgwick D., Norberg P., Peacock J. A., Percival W., Peterson B. A., Sutherland W., Taylor K., 2002, MNRAS, 334, 673

Li B., Hellwing W. A., Koyama K., Zhao G.-B., Jennings E., Baugh C. M., 2012, arXiv:astro-ph/1206.4317

Li I. H., Glazebrook K., Gilbank D., Balogh M., Bower R., Baldry I., Davies G., Hau G., McCarthy P., 2011, MNRAS, 411, 1869

Lilly S. J., Cowie L. L., Gardner J. P., 1991, ApJ, 369, 79

Limber D. N., 1953, ApJ, 117, 134

Limber D. N., Mathews W. G., 1960, ApJ, 132, 286

Lin L., Cooper M. C., Jian H.-Y., Koo D. C., Patton D. R., Yan R., Willmer C. N. A., Coil A. L., Chiueh T., Croton D. J., Gerke B. F., Lotz J., Guhathakurta P., Newman J. A., 2010, ApJ, 718, 1158

Lin L., Dickinson M., Jian H.-Y., Merson A. I., Baugh C. M., Scott D., Foucaud S., Wang W.-H., Yan C.-H., Yan H.-J., Cheng Y.-W., Guo Y., Helly J., Kirsten F., Koo D. C.,

Lagos C. d. P., Meger N., Messias H., Pope A., Simard L., Grogin N. A., Wang S.-Y., 2012, ApJ, 756, 71

Lin L., Patton D. R., Koo D. C., Casteels K., Conselice C. J., Faber S. M., Lotz J., Willmer C. N. A., Hsieh B. C., Chiueh T., Newman J. A., Novak G. S., Weiner B. J., Cooper M. C., 2008, ApJ, 681, 232

Loh E. D., Spillar E. J., 1986, ApJ, 303, 154

Madau P., 1995, ApJ, 441, 18

Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106

Maddox S. J., Efstathiou G., Sutherland W. J., Loveday J., 1990, MNRAS, 242, 43P

Madgwick D. S., Lahav O., Baldry I. K., Baugh C. M., Bland-Hawthorn J., Bridges T., Cannon R., Cole S., Colless M., Collins C., Couch W., Dalton G., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Frenk C. S., Glazebrook K., Jackson C., Lewis I., Lumsden S., Maddox S., Norberg P., Peacock J. A., Peterson B. A., Sutherland W., Taylor K., 2002, MNRAS, 333, 133

Magorrian J., Tremaine S., Richstone D., Bender R., Bower G., Dressler A., Faber S. M., Gebhardt K., Green R., Grillmair C., Kormendy J., Lauer T., 1998, AJ, 115, 2285

Magoulas C., Springob C. M., Colless M., Jones D. H., Campbell L. A., Lucey J. R., Mould J., Jarrett T., Merson A., Brough S., 2012, arXiv:astro-ph/1206.0385

Malbon R. K., Baugh C. M., Frenk C. S., Lacey C. G., 2007, MNRAS, 382, 1394

Maraston C., 2005, MNRAS, 362, 799

Marinoni C., Davis M., Newman J. A., Coil A. L., 2002, ApJ, 580, 122

Martin C. L., 1999, ApJ, 513, 156

Martin C. L., Kennicutt Jr. R. C., 2001, ApJ, 555, 301

Martínez H. J., O'Mill A. L., Lambas D. G., 2006, MNRAS, 372, 253

McBride C., Berlind A., Scoccimarro R., Wechsler R., Busha M., Gardner J., van den Bosch F., 2009, in Bulletin of the American Astronomical Society, Vol. 41, American Astronomical Society Meeting Abstracts 213, p. 425.06

McCarthy P. J., 2004, ARA&A, 42, 477

McCracken H. J., Capak P., Salvato M., Aussel H., Thompson D., Daddi E., Sanders D. B., Kneib J.-P., Willott C. J., Mancini C., Renzini A., Cook R., Le Fèvre O., Ilbert O., Kartaltepe J., Koekemoer A. M., Mellier Y., Murayama T., Scoville N. Z., Shioya Y., Tanaguchi Y., 2010, ApJ, 708, 202

McCracken H. J., Radovich M., Bertin E., Mellier Y., Dantel-Fort M., Le Fèvre O., Cuillandre J. C., Gwyn S., Foucaud S., Zamorani G., 2003, A&A, 410, 17

McCrady N., Gilbert A. M., Graham J. R., 2003, ApJ, 596, 240

Meiksin A., 2006, MNRAS, 365, 807

Merchán M., Zandivarez A., 2002, MNRAS, 335, 216

Mercurio A., Haines C. P., Gargiulo A., La Barbera F., Merluzzi P., Busarello G., 2010, arXiv:astro-ph/1006.5001

Merritt D., Navarro J. F., Ludlow A., Jenkins A., 2005, ApJ, 624, L85

Metcalfe N., Shanks T., Weilbacher P. M., McCracken H. J., Fong R., Thompson D., 2006, MNRAS, 370, 1257

Miller C. J., Nichol R. C., Reichart D., Wechsler R. H., Evrard A. E., Annis J., McKay T. A., Bahcall N. A., Bernardi M., Boehringer H., Connolly A. J., Goto T., Kniazev A., Lamb D., Postman M., Schneider D. P., Sheth R. K., Voges W., 2005, AJ, 130, 968

Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution

Mohr J. J., Armstrong R., Bertin E., Daues G. E., Desai S., Gower M., Gruendl R., Hanlon W., Kuropatkin N., Lin H., Marriner J., Petravick D., Sevilla I., Swanson M., Tomashek T., Tucker D., Yanny B., the Dark Energy Survey Collaboration, 2012, arXiv:astro-ph/1207.3189

Moles M., Benítez N., Aguerri J. A. L., Alfaro E. J., Broadhurst T., Cabrera-Caño J., Castander F. J., Cepa J., Cerviño M., Cristóbal-Hornillos D., Fernández-Soto A., González Delgado R. M., Infante L., Márquez I., Martínez V. J., Masegosa J., del Olmo A., Perea J., Prada F., Quintana J. M., Sánchez S. F., 2008, AJ, 136, 1325

Murphy D. N. A., Eke V. R., Frenk C. S., 2011, MNRAS, 413, 2288

Murphy D. N. A., Geach J. E., Bower R. G., 2012, MNRAS, 420, 1861

Nagashima M., Lacey C. G., Baugh C. M., Frenk C. S., Cole S., 2005, MNRAS, 358, 1247

Navarro J. F., Frenk C. S., White S. D. M., 1995, MNRAS, 275, 720

—, 1996, ApJ, 462, 563

—, 1997, ApJ, 490, 493

Navarro J. F., Hayashi E., Power C., Jenkins A. R., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T. R., 2004, MNRAS, 349, 1039

Navarro J. F., White S. D. M., 1993, MNRAS, 265, 271

—, 1994, MNRAS, 267, 401

Newman J. A., Cooper M. C., Davis M., Faber S. M., Coil A. L., Guhathakurta P., Koo D. C., Phillips A. C., Conroy C., Dutton A. A., Finkbeiner D. P., Gerke B. F., Rosario D. J., Weiner B. J., Willmer C. N. A., Yan R., Harker J. J., Kassin S. A., Konidaris N. P., Lai K., Madgwick D. S., Noeske K. G., Wirth G. D., Connolly A. J., Kaiser N., Kirby E. N., Lemaux B. C., Lin L., Lotz J. M., Luppino G. A., Marinoni C., Matthews D. J., Metevier A., Schiavon R. P., 2012, arXiv:astro-ph/1203.3192

Norberg P., Baugh C. M., Hawkins E., Maddox S., Madgwick D., Lahav O., Cole S., Frenk C. S., Baldry I., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Dalton G., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Glazebrook K., Jackson C., Lewis I., Lumsden S., Peacock J. A., Peterson B. A., Sutherland W., Taylor K., 2002, MNRAS, 332, 827

Norberg P., Baugh C. M., Hawkins E., Maddox S., Peacock J. A., Cole S., Frenk C. S., Bland-Hawthorn J., Bridges T., Cannon R., Colless M., Collins C., Couch W., Dalton G., De Propris R., Driver S. P., Efstathiou G., Ellis R. S., Glazebrook K., Jackson C., Lahav O., Lewis I., Lumsden S., Madgwick D., Peterson B. A., Sutherland W., Taylor K., 2001, MNRAS, 328, 64

Ocvirk P., Pichon C., Teyssier R., 2008, MNRAS, 390, 1326

Okamoto T., Gao L., Theuns T., 2008, MNRAS, 390, 920

Oke J. B., Gunn J. E., 1983, ApJ, 266, 713

Oort J. H., 1932, Bull. Astron. Inst. Netherlands, 6, 249

Orsi A., Baugh C. M., Lacey C. G., Cimatti A., Wang Y., Zamorani G., 2010, MNRAS, 405, 1006

Overzier R. A., Guo Q., Kauffmann G., De Lucia G., Bouwens R., Lemson G., 2009, MNRAS, 394, 577

Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557

Parra R., Conway J. E., Diamond P. J., Thrall H., Lonsdale C. J., Lonsdale C. J., Smith H. E., 2007, ApJ, 659, 314

Paumard T., Genzel R., Martins F., Nayakshin S., Beloborodov A. M., Levin Y., Trippe S., Eisenhauer F., Ott T., Gillessen S., Abuter R., Cuadra J., Alexander T., Sternberg A., 2006, ApJ, 643, 1011

Peacock J. A., 1997, MNRAS, 284, 885

—, 1999, Cosmological Physics

Peacock J. A., Heavens A. F., 1990, MNRAS, 243, 133

Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

Peebles P. J. E., 1980, The large-scale structure of the universe

Perlmutter S., Aldering G., Goldhaber G., Knop R. A., Nugent P., Castro P. G., Deustua S., Fabbro S., Goobar A., Groom D. E., Hook I. M., Kim A. G., Kim M. Y., Lee J. C., Nunes N. J., Pain R., Pennypacker C. R., Quimby R., Lidman C., Ellis R. S., Irwin M., McMahon R. G., Ruiz-Lapuente P., Walton N., Schaefer B., Boyle B. J., Filippenko A. V., Matheson T., Fruchter A. S., Panagia N., Newberg H. J. M., Couch W. J., Supernova Cosmology Project, 1999, ApJ, 517, 565

Peterson J. R., Kahn S. M., Paerels F. B. S., Kaastra J. S., Tamura T., Bleeker J. A. M., Ferrigno C., Jernigan J. G., 2003, ApJ, 590, 207

Pettini M., Rix S. A., Steidel C. C., Adelberger K. L., Hunt M. P., Shapley A. E., 2002, ApJ, 569, 742

Poggianti B. M., 1997, A&AS, 122, 399

Poggianti B. M., von der Linden A., De Lucia G., Desai V., Simard L., Halliday C., Aragón-Salamanca A., Bower R., Varela J., Best P., Clowe D. I., Dalcanton J., Jablonka P., Milvang-Jensen B., Pello R., Rudnick G., Saglia R., White S. D. M., Zaritsky D., 2006, ApJ, 642, 188

Power C., Baugh C. M., Lacey C. G., 2010, MNRAS, 406, 43

Prada F., Klypin A. A., Simonneau E., Betancort-Rijo J., Patiri S., Gottlöber S., Sanchez-Conde M. A., 2006, ApJ, 645, 1001

Press W. H., Schechter P., 1974, ApJ, 187, 425

Quadri R., Marchesini D., van Dokkum P., Gawiser E., Franx M., Lira P., Rudnick G., Urry C. M., Maza J., Kriek M., Barrientos L. F., Blanc G. A., Castander F. J., Christlein D., Coppi P. S., Hall P. B., Herrera D., Infante L., Taylor E. N., Treister E., Willis J. P., 2007, AJ, 134, 1103

Reddy N. A., Erb D. K., Steidel C. C., Shapley A. E., Adelberger K. L., Pettini M., 2005, ApJ, 633, 748

Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541

Rieke G. H., Loken K., Rieke M. J., Tamblyn P., 1993, ApJ, 412, 99

Riess A. G., Filippenko A. V., Challis P., Clocchiatti A., Diercks A., Garnavich P. M., Gilliland R. L., Hogan C. J., Jha S., Kirshner R. P., Leibundgut B., Phillips M. M., Reiss D., Schmidt B. P., Schommer R. A., Smith R. C., Spyromilio J., Stubbs C., Suntzeff N. B., Tonry J., 1998, AJ, 116, 1009

Robotham A., Phillipps S., de Propris R., 2010, MNRAS, 403, 1812

Robotham A. S. G., Norberg P., Driver S. P., Baldry I. K., Bamford S. P., Hopkins A. M., Liske J., Loveday J., Merson A., Peacock J. A., Brough S., Cameron E., Conselice C. J., Croom S. M., Frenk C. S., Gunawardhana M., Hill D. T., Jones D. H., Kelvin L. S., Kuijken K., Nichol R. C., Parkinson H. R., Pimbblet K. A., Phillipps S., Popescu C. C., Prescott M., Sharp R. G., Sutherland W. J., Taylor E. N., Thomas D., Tuffs R. J., van Kampen E., Wijesinghe D., 2011, MNRAS, 416, 2640

Roche N., Eales S. A., 1999, MNRAS, 307, 703

Ross A. J., Percival W. J., Brunner R. J., 2010, MNRAS, 407, 420

Ross N. P., da Ângela J., Shanks T., Wake D. A., Cannon R. D., Edge A. C., Nichol R. C., Outram P. J., Colless M., Couch W. J., Croom S. M., de Propris R., Drinkwater M. J., Eisenstein D. J., Loveday J., Pimbblet K. A., Roseboom I. G., Schneider D. P., Sharp R. G., Weilbacher P. M., 2007, MNRAS, 381, 573

Rovilos E., Burwitz V., Szokoly G., Hasinger G., Egami E., Bouché N., Berta S., Salvato M., Lutz D., Genzel R., 2009, A&A, 507, 195

Rowan-Robinson M., Babbedge T., Oliver S., Trichas M., Berta S., Lonsdale C., Smith G., Shupe D., Surace J., Arnouts S., Ilbert O., Le Févre O., Afonso-Luis A., Perez-Fournon I., Hatziminaoglou E., Polletta M., Farrah D., Vaccari M., 2008, MNRAS, 386, 697

Rubin V. C., Ford W. K. J., . Thonnard N., 1980, ApJ, 238, 471

Rubin V. C., Thonnard N., Ford Jr. W. K., 1978, ApJ, 225, L107

Salpeter E. E., 1955, ApJ, 121, 161

Sánchez A. G., Crocce M., Cabré A., Baugh C. M., Gaztañaga E., 2009, MNRAS, 400, 1643

Sánchez A. G., Scóccola C. G., Ross A. J., Percival W., Manera M., Montesano F., Mazzalay X., Cuesta A. J., Eisenstein D. J., Kazin E., McBride C. K., Mehta K., Montero-Dorta A. D., Padmanabhan N., Prada F., Rubiño-Martín J. A., Tojeiro R., Xu X., Magaña M. V., Aubourg E., Bahcall N. A., Bailey S., Bizyaev D., Bolton A. S., Brewington H., Brinkmann J., Brownstein J. R., Gott J. R., Hamilton J. C., Ho S., Honscheid K., Labatie A., Malanushenko E., Malanushenko V., Maraston C., Muna D., Nichol R. C., Oravetz D., Pan K., Ross N. P., Roe N. A., Reid B. A., Schlegel D. J., Shelden A., Schneider D. P., Simmons A., Skibba R., Snedden S., Thomas D., Tinker J., Wake D. A., Weaver B. A., Weinberg D. H., White M., Zehavi I., Zhao G., 2012, MNRAS, 425, 415

Saracco P., Giallongo E., Cristiani S., D'Odorico S., Fontana A., Iovino A., Poli F., Vanzella E., 2001, A&A, 375, 1

Saunders W., Frenk C., Rowan-Robinson M., Lawrence A., Efstathiou G., 1991, Nature, 349, 32

Savage B. D., Mathis J. S., 1979, ARA&A, 17, 73

Schaye J., 2004, ApJ, 609, 667

Schaye J., Dalla Vecchia C., Booth C. M., Wiersma R. P. C., Theuns T., Haas M. R., Bertone S., Duffy A. R., McCarthy I. G., van de Voort F., 2010, MNRAS, 402, 1536

Schechter P., 1976, ApJ, 203, 297

Schlegel D., White M., Eisenstein D., 2009, in ArXiv Astrophysics e-prints, Vol. 2010, astro2010: The Astronomy and Astrophysics Decadal Survey, p. 314

Schmidt M., 1959, ApJ, 129, 243

Seljak U., 2000, MNRAS, 318, 203

Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437

Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS, 343, 978

Shu F. H., 1973, in IAU Symposium, Vol. 52, Interstellar Dust and Related Topics, Greenberg J. M., van de Hulst H. C., eds., p. 257

Silk J., 1977, ApJ, 211, 638

Silk J., Norman C., 2009, ApJ, 700, 262

Silk J., Rees M. J., 1998, A&A, 331, L1

Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, 103

Simon P., 2007, A&A, 473, 711

Skibba R. A., Bamford S. P., Nichol R. C., Lintott C. J., Andreescu D., Edmondson E. M., Murray P., Raddick M. J., Schawinski K., Slosar A., Szalay A. S., Thomas D., Vandenberg J., 2009, MNRAS, 399, 966

Smail I., Chapman S. C., Ivison R. J., Blain A. W., Takata T., Heckman T. M., Dunlop J. S., Sekiguchi K., 2003, MNRAS, 342, 1185

Smoot G. F., Bennett C. L., Kogut A., Wright E. L., Aymon J., Boggess N. W., Cheng E. S., de Amici G., Gulkis S., Hauser M. G., Hinshaw G., Jackson P. D., Janssen M., Kaita E., Kelsall T., Keegstra P., Lineweaver C., Loewenstein K., Lubin P., Mather J., Meyer S. S., Moseley S. H., Murdock T., Rokke L., Silverberg R. F., Tenorio L., Weiss R., Wilkinson D. T., 1992, ApJ, 396, L1

Sobral D., Best P. N., Smail I., Geach J. E., Cirasuolo M., Garn T., Dalton G. B., 2011, MNRAS, 411, 675

Song J., Mohr J. J., Barkhouse W. A., Warren M. S., Dolag K., Rude C., 2012, ApJ, 747, 58

Sousbie T., Courtois H., Bryan G., Devriendt J., 2008, ApJ, 678, 569

Spergel D. N., Bean R., Doré O., Nolta M. R., Bennett C. L., Dunkley J., Hinshaw G., Jarosik N., Komatsu E., Page L., Peiris H. V., Verde L., Halpern M., Hill R. S., Kogut A., Limon M., Meyer S. S., Odegard N., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2007, ApJS, 170, 377

Spergel D. N., Verde L., Peiris H. V., Komatsu E., Nolta M. R., Bennett C. L., Halpern M., Hinshaw G., Jarosik N., Kogut A., Limon M., Meyer S. S., Page L., Tucker G. S., Weiland J. L., Wollack E., Wright E. L., 2003, ApJS, 148, 175

Springel V., 2005, MNRAS, 364, 1105

—, 2010, ARA&A, 48, 391

Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005, Nature, 435, 629

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Steidel C. C., Adelberger K. L., Shapley A. E., Pettini M., Dickinson M., Giavalisco M., 2003, ApJ, 592, 728

Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17

Steidel C. C., Shapley A. E., Pettini M., Adelberger K. L., Erb D. K., Reddy N. A., Hunt M. P., 2004, ApJ, 604, 534

Stolte A., Brandner W., Grebel E. K., Lenzen R., Lagrange A.-M., 2005, ApJ, 628, L113

Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253

Tamura T., Kaastra J. S., Peterson J. R., Paerels F. B. S., Mittaz J. P. D., Trudolyubov S. P., Stewart G., Fabian A. C., Mushotzky R. F., Lumb D. H., Ikebe Y., 2001, A&A, 365, L87

Tan J. C., 2000, ApJ, 536, 173

Tanaka M., Finoguenov A., Lilly S. J., Bolzonella M., Carollo C. M., Contini T., Iovino A., Kneib J.-P., Lamareille F., Le Fevre O., Mainieri V., Presotto V., Renzini A., Scodeggio M., Silverman J. D., Zamorani G., Bardelli S., Bongiorno A., Caputi K., Cucciati O., de la Torre S., de Ravel L., Franzetti P., Garilli B., Kampczyk P., Knobel C., Kovač K., Le Borgne J.-F., Le Brun V., López-Sanjuan C., Maier C., Mignoli M., Pello R., Peng Y., Perez-Montero E., Tasca L., Tresse L., Vergani D., Zucca E., Barnes L., Bordoloi R., Cappi A., Cimatti A., Coppa G., Koekemoer A. M., McCracken H. J., Moresco M., Nair P., Oesch P., Pozzetti L., Welikala N., 2012, PASJ, 64, 22

Tegmark M., Blanton M. R., Strauss M. A., Hoyle F., Schlegel D., Scoccimarro R., Vogeley M. S., Weinberg D. H., Zehavi I., Berlind A., Budavari T., Connolly A., Eisenstein D. J., Finkbeiner D., Frieman J. A., Gunn J. E., Hamilton A. J. S., Hui L., Jain B., Johnston D., Kent S., Lin H., Nakajima R., Nichol R. C., Ostriker J. P., Pope A., Scranton R., Seljak U., Sheth R. K., Stebbins A., Szalay A. S., Szapudi I., Verde L., Xu Y., Annis J., Bahcall N. A., Brinkmann J., Burles S., Castander F. J., Csabai I., Loveday J., Doi M., Fukugita M., Gott III J. R., Hennessy G., Hogg D. W., Ivezić Ž., Knapp G. R., Lamb D. Q., Lee B. C., Lupton R. H., McKay T. A., Kunszt P., Munn J. A., O'Connell L., Peoples J., Pier J. R., Richmond M., Rockosi C., Schneider D. P., Stoughton C., Tucker D. L., Vanden Berk D. E., Yanny B., York D. G., SDSS Collaboration, 2004, ApJ, 606, 702

Tegmark M., Eisenstein D. J., Strauss M. A., Weinberg D. H., Blanton M. R., Frieman J. A., Fukugita M., Gunn J. E., Hamilton A. J. S., Knapp G. R., Nichol R. C., Ostriker J. P., Padmanabhan N., Percival W. J., Schlegel D. J., Schneider D. P., Scoccimarro R., Seljak U., Seo H.-J., Swanson M., Szalay A. S., Vogeley M. S., Yoo J., Zehavi I., Abazajian K., Anderson S. F., Annis J., Bahcall N. A., Bassett B., Berlind A., Brinkmann J., Budavari T., Castander F., Connolly A., Csabai I., Doi M., Finkbeiner D. P., Gillespie B., Glazebrook K., Hennessy G. S., Hogg D. W., Ivezić Ž., Jain B., Johnston D., Kent S., Lamb D. Q., Lee B. C., Lin H., Loveday J., Lupton R. H., Munn J. A., Pan K., Park C., Peoples J., Pier J. R., Pope A., Richmond M., Rockosi C., Scranton R., Sheth R. K., Stebbins A., Stoughton C., Szapudi I., Tucker D. L., vanden Berk D. E., Yanny B., York D. G., 2006, Phys. Rev. D, 74, 123507

Tinsley B. M., 1980, Fund. Cosmic Phys., 5, 287

Tormen G., 1997, MNRAS, 290, 411

Tran K.-V. H., Moustakas J., Gonzalez A. H., Bai L., Zaritsky D., Kautsch S. J., 2008, ApJ, 683, L17

Tully R. B., Fisher J. R., 1977, A&A, 54, 661

Vale A., Ostriker J. P., 2004, MNRAS, 353, 189

Valotto C. A., Nicotra M. A., Muriel H., Lambas D. G., 1997, ApJ, 479, 90

van der Wel A., Bell E. F., Holden B. P., Skibba R. A., Rix H.-W., 2010, ApJ, 714, 1779

van Dokkum P. G., 2008, ApJ, 674, 29

van Dokkum P. G., Förster Schreiber N. M., Franx M., Daddi E., Illingworth G. D., Labbé I., Moorwood A., Rix H.-W., Röttgering H., Rudnick G., van der Wel A., van der Werf P., van Starkenburg L., 2003, ApJ, 587, L83

Vandame B., Olsen L. F., Jorgensen H. E., Groenewegen M. A. T., Schirmer M., Arnouts S., Benoist C., da Costa L., Mignani R. P., Rite' C., Slijkhuis R., Hatziminaoglou E., Hook R., Madejsky R., Wicenec A., 2001, ArXiv Astrophysics e-prints

Wake D. A., Sheth R. K., Nichol R. C., Baugh C. M., Bland-Hawthorn J., Colless M., Couch W. J., Croom S. M., de Propris R., Drinkwater M. J., Edge A. C., Loveday J., Lam T. Y., Pimbblet K. A., Roseboom I. G., Ross N. P., Schneider D. P., Shanks T., Sharp R. G., 2008, MNRAS, 387, 1045

Wang Y., Yang X., Mo H. J., van den Bosch F. C., Weinmann S. M., Chu Y., 2008, ApJ, 687, 919

White S. D. M., 1984, ApJ, 286, 38

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Willmer C. N. A., Faber S. M., Koo D. C., Weiner B. J., Newman J. A., Coil A. L., Connolly A. J., Conroy C., Cooper M. C., Davis M., Finkbeiner D. P., Gerke B. F., Guhathakurta P., Harker J., Kaiser N., Kassin S., Konidaris N. P., Lin L., Luppino G., Madgwick D. S., Noeske K. G., Phillips A. C., Yan R., 2006, ApJ, 647, 853

Wolf C., Meisenheimer K., Rix H.-W., Borch A., Dye S., Kleinheinrich M., 2003, A&A, 401, 73

Wolf C., Meisenheimer K., Röser H.-J., Beckwith S. V. W., Chaffee Jr. F. H., Fried J., Hippelein H., Huang J.-S., Kümmel M., von Kuhlmann B., Maier C., Phleps S., Rix H.-W., Thommes E., Thompson D., 2001, A&A, 365, 681

Wong T., Blitz L., 2002, ApJ, 569, 157

Wu K. K. S., Fabian A. C., Nulsen P. E. J., 2000, MNRAS, 318, 889

Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., 2005, MNRAS, 356, 1293

Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153

York D. G., Adelman J., Anderson Jr. J. E., Anderson S. F., Annis J., Bahcall N. A., Bakken J. A., Barkhouser R., Bastian S., Berman E., Boroski W. N., Bracker S., Briegel C., Briggs J. W., Brinkmann J., Brunner R., Burles S., Carey L., Carr M. A., Castander F. J., Chen B., Colestock P. L., Connolly A. J., Crocker J. H., Csabai I., Czarapata P. C., Davis J. E., Doi M., Dombeck T., Eisenstein D., Ellman N., Elms B. R., Evans M. L., Fan X., Federwitz G. R., Fiscelli L., Friedman S., Frieman J. A., Fukugita M., Gillespie B., Gunn J. E., Gurbani V. K., de Haas E., Haldeman M., Harris F. H., Hayes J., Heckman T. M., Hennessy G. S., Hindsley R. B., Holm S., Holmgren D. J., Huang C.-h., Hull C., Husby D., Ichikawa S.-I., Ichikawa T., Ivezić Ž., Kent S., Kim R. S. J., Kinney E., Klaene M., Kleinman A. N., Kleinman S., Knapp G. R., Korienek J., Kron R. G., Kunszt P. Z., Lamb D. Q., Lee B., Leger R. F., Limmongkol S., Lindenmeyer C., Long D. C., Loomis C., Loveday J., Lucinio R., Lupton R. H., MacKinnon B., Mannery E. J., Mantsch P. M., Margon B., McGehee P., McKay T. A., Meiksin A., Merelli A., Monet D. G., Munn J. A., Narayanan V. K., Nash T., Neilsen E., Neswold R., Newberg H. J., Nichol R. C., Nicinski T., Nonino M., Okada N., Okamura S., Ostriker J. P., Owen R., Pauls A. G., Peoples J., Peterson R. L., Petravick D., Pier J. R., Pope A., Pordes R., Prosapio A., Rechenmacher R., Quinn T. R., Richards G. T., Richmond M. W., Rivetta C. H., Rockosi C. M., Ruthmansdorfer K., Sandford D., Schlegel D. J., Schneider D. P., Sekiguchi M., Sergey G., Shimasaku K., Siegmund W. A., Smee S., Smith J. A., Snedden S., Stone R., Stoughton C., Strauss M. A., Stubbs C., SubbaRao M., Szalay A. S., Szapudi I., Szokoly G. P., Thakar A. R., Tremonti C., Tucker D. L.,

Uomoto A., Vanden Berk D., Vogeley M. S., Waddell P., Wang S.-i., Watanabe M., Weinberg D. H., Yanny B., Yasuda N., SDSS Collaboration, 2000, AJ, 120, 1579

Yoshida N., Stoehr F., Springel V., White S. D. M., 2002, MNRAS, 335, 762

Zavala J., Okamoto T., Frenk C. S., 2008, MNRAS, 387, 364

Zehavi I., Blanton M. R., Frieman J. A., Weinberg D. H., Mo H. J., Strauss M. A., Anderson S. F., Annis J., Bahcall N. A., Bernardi M., Briggs J. W., Brinkmann J., Burles S., Carey L., Castander F. J., Connolly A. J., Csabai I., Dalcanton J. J., Dodelson S., Doi M., Eisenstein D., Evans M. L., Finkbeiner D. P., Friedman S., Fukugita M., Gunn J. E., Hennessy G. S., Hindsley R. B., Ivezić Ž., Kent S., Knapp G. R., Kron R., Kunszt P., Lamb D. Q., Leger R. F., Long D. C., Loveday J., Lupton R. H., McKay T., Meiksin A., Merrelli A., Munn J. A., Narayanan V., Newcomb M., Nichol R. C., Owen R., Peoples J., Pope A., Rockosi C. M., Schlegel D., Schneider D. P., Scoccimarro R., Sheth R. K., Siegmund W., Smee S., Snir Y., Stebbins A., Stoughton C., SubbaRao M., Szalay A. S., Szapudi I., Tegmark M., Tucker D. L., Uomoto A., Vanden Berk D., Vogeley M. S., Waddell P., Yanny B., York D. G., 2002, ApJ, 571, 172

Zehavi I., Zheng Z., Weinberg D. H., Blanton M. R., Bahcall N. A., Berlind A. A., Brinkmann J., Frieman J. A., Gunn J. E., Lupton R. H., Nichol R. C., Percival W. J., Schneider D. P., Skibba R. A., Strauss M. A., Tegmark M., York D. G., 2011, ApJ, 736, 59

Zehavi I., Zheng Z., Weinberg D. H., Frieman J. A., Berlind A. A., Blanton M. R., Scoccimarro R., Sheth R. K., Strauss M. A., Kayo I., Suto Y., Fukugita M., Nakamura O., Bahcall N. A., Brinkmann J., Gunn J. E., Hennessy G. S., Ivezić Ž., Knapp G. R., Loveday J., Meiksin A., Schlegel D. J., Schneider D. P., Szapudi I., Tegmark M., Vogeley M. S., York D. G., SDSS Collaboration, 2005, ApJ, 630, 1

Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

Zwicky F., 1933, Helvetica Physica Acta, 6, 110